Strategies of viral multi-functional regulator proteins Adeno-associated virus rep

Bardelli, Martino

Awarding institution:
King's College London

Strategies of Viral Multi-functional Regulator Proteins: Adeno-Associated Virus Rep

Martino Bardelli

This thesis is submitted for the degree of Doctor of Philosophy
at King's College London

Department of Infectious Diseases
Faculty of Life Sciences & Medicine

- September 2015 -

A nonno Giovanni e nonno Marino

ABSTRACT

Adeno-associated virus (AAV) is a small non-pathogenic human DNA parvovirus. The AAV life cycle, which includes transcriptional regulation, DNA replication, assembly and site-specific integration, is orchestrated by AAV's four Rep proteins. Structurally, these proteins share a AAA+ domain characteristic of the SF3 family of helicases, with the larger Rep68 and Rep78 additionally containing an N-terminal origin-binding domain (OBD) that specifically binds and nicks DNA. The combination of these domains is the basis for the remarkable multi-functionality displayed by Rep68 and Rep78. To date, structural studies of Rep68 and Rep78 have been limited by the tendency of these proteins to aggregate when purified. Here, we describe a fully functional Rep mutant that does not aggregate even at high concentrations and use this mutant to investigate the structural requirements for Rep functions. We demonstrate that one of the determinants regulating the oligomerisation of the Rep proteins lies in the linker connecting the helicase domain and OBD. We also identify a series of key residues at the interface between Rep monomers and show that mutating them has drastic effects both on the oligomerisation and functionality of the Rep proteins. Importantly, these oligomerisation-deficient mutants do not support the AAV life-cycle and fail to bind DNA efficiently, an important Rep function necessary for DNA nicking, transcriptional regulation, viral DNA replication and site-specific integration.

Finally, understanding the molecular details of Rep and its functions will contribute to the development of new AAV-based vectors that exploit the Rep-mediated integration mechanism and potentially have a lower risk of insertional mutagenesis than retroviral vectors. In the last chapter, we describe an AAV vector for testing the safety and feasibility of AAV-mediated targeted gene addition in induced pluripotent stem (iPS) cells within the therapeutic context of SCID-X1, an immunodeficiency caused by mutations in the common gamma chain gene.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS

First and foremost I want to thank my supervisors, Els Henckaerts and Michael Linden, for the many opportunities they have given me and their guidance throughout these years. There have been challenging times during my PhD, but their support helped me to go through the frustrations of scientific research and for that I am very thankful, particularly to Els who has been incredibly supportive during the last intense year. When I first joined the lab as a master student 5 years ago their passion for science, and AAV biology in particular, inspired me to pursue a PhD. Today, their passion is unchanged, and discussing science with them is always an eye-opener.

I would also like to thank all the members of the lab for their help, feedback and discussion, and for making my time in the lab a thoroughly enjoyable one. Starting with Nathalie, who welcomed me to the lab and taught me my first AAV assays, to Leti, who has now left the lab but has been a great colleague and shared with me her passion for science and TrwC; and all the current members of the lab. André, who is a great person to work with and never ceases to surprise with his witty remarks. To Sarah, whose determination and drive have been an inspiration. Julie, thank you for being such a joy to have around, even when grumpy, and for your great insights into AAV vectors. To Nuria, it has been a pleasure to see you grow as a scientist and contribute to the well-being of the lab. To the troublesome “iPSCs ladies”, Zoe and Stina, it was fun to see you learn how to generate iPSCs and differentiate them, going through the beautiful struggles intrinsic to iPSCs cultures.

I am particularly thankful to my PhD thesis committee members, Mike Malim, Stuart Neil, Chad Swanson and Adrian Thrasher for feedback and suggestions throughout my PhD, and for helping me stay on track with my progress. Stuart and Chad, in particular, were always willing to discuss scientific and technical problems, and their insights have been invaluable.

I also want to thank everyone in the program of infection and immunity for making it such a stimulating environment to work in. The scientific (and less scientific) discussions at the bench, at seminars, in tissue culture, in the corridors and at the pub have been thoroughly enriching. A particular thank you goes to Rafa, Tonya, Torsten, Julia, Johnny, James, Terry, Val, Suzy and Susi, Tosh, Greg, Oliver, Fernanda, Luise and Eva, Lauren, Caroline, Shereen, Robin, Debby, Shetal, Gina, Tom, and many more. A particular mention goes to Rui, who has been somewhere half way between a big brother and a mentor, has never held back from coming to my rescue, and has contributed greatly to my scientific development. His comments on the thesis, as always, were thought-provoking as expected from a true virologist.

On a more personal level, but surely not of less scientific relevance, I want to thank Anna, who has been beyond immensely supportive during these years, and even more so during these last months whilst writing up. Her scientific expertise, combined with her drive and passion has enriched my life on many levels.

Finally, last but not least I want to thank my family, Michi, Lollo e Nora for their unconditional love and support, and for teaching me since I was kid how to be part of a great team! So che Londra non è proprio a due passi per voi, ma malgrado la lontananza fisica non avete mai mancato di starmi vicino e di incoraggiarmi a intraprendere decisioni che non vi andavano a favore. Grazie infinitamente, non so come farei senza di voi..

ABBREVIATIONS

293T	Human embryonic kidney 293 T-antigen cells
AA	Amino acid residue
AAA+	ATPases associated with diverse cellular activities
AAP	Assembly-activating protein
AAV	Adeno-associated virus
AAVS1	AAV integration site 1
Ad	Adenovirus
ADA-SCID	Adenosine deaminase SCID
ADP	Adenosine diphosphate
ATM	Ataxia telangiectasia mutated
ATP	Adenosine triphosphate
bp	Base pairs
BSA	Bovine serum albumin
CGD	Chronic granulomatous disease
CLIC/GEEC	Clathrin-independent carriers/GPI-enriched endocytic compartment
CMV	Cytomegalovirus
CMVie	Cytomegalovirus immediate-early
CPE	Cytopathogenic effect
CRISPR	Clustered regularly interspaced short palindromic repeats
Cys, C	Cysteine
DBP	DNA binding protein of adenovirus (product of the E2A gene)
ddH_2O	Double distilled water
DDR	DNA damage response
DMEM	Dulbecco's Modified Eagle Medium
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
DSB	Double strand break
dsDNA	Double-stranded DNA
E1	E1 protein of papilloma virus
EDTA	1-(4-Aminobenzyl)ethlenediamine-N,N,N',N'-tetraacetic acid

EM	Electron microscopy
ESCs	Embryonic stem cells
FBS	Fetal bovine serum
GFP	Green fluorescent protein
GMP	Good manufacturing practice
GPI	Glycosylphosphatidylinositol
HeLa	Henrietta Lacks (Human epitheloid carcinoma cells)
His, H	Histidine
HLA	Human leukocyte antigen
HMG1	High mobility group protein 1
HR	Homologous recombination
HRP	Horseradish Peroxidase
HSCs	Hematopoietic stem cells
HSPG	Heparan sulphate proteoglycan
HSV-1	Herpes simplex virus 1
HUH	His-hydrophobic-His
IL-2	Interleukin 2
IL2RG	IL-2 receptor common gamma chain
Ile, I	Isoleucine
iPSCs	Induced pluripotent stem cells
ITR	Inverted terminal repeat
LB	Lysogeny broth
LTag	Large T antigen of SV40
Lys, K	Lysine
LTR	Long terminal repeat
MBP	Maltose-binding protein
Mbs85	Myosin-binding subunit 85
MCM	Minichromosome maintenance complex
MOI	Multiplicity of infection
MRN	Mre11, Rad50, Nbs1 complex
mRNA	Messenger RNA
MVM	Minute virus of mice

NEB	New England Biolabs
NHEJ	Non-homologous end-joining
NTP	Nucleoside triphosphate
OBD	Origin binding site
OD	Oligomerisation domain
ORF	Open reading frame
ori	Origin of replication
PAGE	Poly-acrylamide gel electrophoresis
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
PCNA	Proliferating nuclear antigen
PEI	Polyethylinimine
PFA	Paraformaldehyde
Phe, F	Phenylalanine
PIDs	Primary immunodeficiencies
PLA_2	Phospholipase A_2
POLD	DNA polymerase delta
PPP1R12C	Protein phosphatase I regulatory inhibitor subunit 12C
PV	Papilloma virus
qPCR	Quantitative real-time PCR
rAAV	Recombinant AAV vectors
RCR	Rolling circle replication
RBS	Rep binding site
RE	Restriction enzyme
Rep*	Rep-C151S (mutation of cysteine 151 to serine)
Rep68*	Rep68-C151S
RFC	Replicating factor C
RGENs	RNA-guided engineered nucleases
RNA	Ribonucleic acid
RT	RT Reverse transcriptase
scAAVs	Self-complementary AAVs
SCID-X1	X-linked severe combined immunodeficiency

Ser, S	Serine
S.e.m.	Standard error of the mean
SF3	Superfamily 3 of helicases
ssDNA	Single-stranded DNA
SV40	Simian virus 40
TAE	Tris-acetate-EDTA
TALENs	Transcription activator-like effector nucleases
TEMED	N,N,N',N'-tetramethylethylenediamine
TRS	Terminal resolution site
Trp, W	Tryptophan
TYLCV	Tomato yellow leaf curl virus
Tyr, Y	Tyrosine
Val, V	Valine
VA RNA	Viral associated RNA
VRs	Variable regions
WAS	Wiskott-Aldrich syndrome
WPRE	Woodchuck Posttranscriptional Regulatory Element
WT	Wild type
ZFNs	Zinc-finger nucleases
Zn	Zinc

INDEX OF FIGURES

INDEX OF TABLES

CHAPTER 1: INTRODUCTION

1.1 ADENO-ASSOCIATED VIRUS

1.1.1 Adeno-associated virus – a historical overview

Adeno-associated virus (AAV) was first discovered 50 years ago as a contaminant of adenovirus stocks (Atchison et al. 1965; Hoggan et al. 1966). It was initially described as a defective small virus-like particle, serologically distinct from adenovirus, which was infectious in humans but required co-infection by adenovirus to replicate. It was soon noticed that other viruses, such as herpes virus, could also support productive AAV replication (Blacklow et al. 1970). Early research, inspired by the molecular biology studies on bacteriophages, focussed on dissecting the AAV genome structure. It was noted that the AAV genome was DNA, that it was packaged as a linear single strand of either plus or minus sense (Crawford et al. 1969; Rose et al. 1969) and that the minus sense DNA was the template for mRNA transcription (Carter et al. 1975). The three structural proteins of AAV were identified soon after and termed VP1, VP2 and VP3 (Johnson et al. 1971; Rose et al. 1971). Further studies into the structure of the AAV DNA revealed that it contained inverted palindromic repeats that could form complex secondary structures (Carter et al. 1972; Gerry et al. 1973; Koczot et al. 1973; Berns and Kelly 1974) and serve as origin of replication, leading to the first models of AAV replication (Denhardt et al. 1976; Straus et al. 1976; Tattersall and Ward 1976; Hauswirth and Berns 1977). Direct sequencing of the AAV termini (Lusby et al. 1980) confirmed the presence of inverted terminal repeats (ITRs) that could form T-shaped hairpins. In the same period, a mechanism for packaging of single stranded progeny DNA into preformed capsids was described (Myers and Carter 1980), and the basic transcriptional map of AAV was defined (Laughlin et al. 1979; Green and Roeder 1980; Marcus et al. 1981). Furthermore, it was observed that AAV could establish latent infection in cells in the absence of helper virus by integrating into the host genome, and that infectious AAV could be rescued from these cells upon infection with adenovirus (Hoggan 1972; Cheung et al. 1980).

Significant progress in the field came from the development of the first AAV molecular clones in the early 1980s (Samulski et al. 1982; Laughlin et al. 1983) and from the sequencing of its genome in 1983 (Srivastava et al. 1983). At this point, genetic analysis of the AAV genome was possible (Hermonat et al. 1984; Tratschin et al. 1984a), and it was confirmed that the ITRs are the origin of AAV replication and act in *cis*. In addition, two large open reading frames (ORF), coding for *trans*-acting functions, were identified within the AAV genome: the left half of the genome was necessary for replication, while the right half coded for the capsid proteins. These discoveries paved the way for the development of AAV as a gene therapy vector.

The first recombinant vectors (rAAVs) were produced by co-transfecting an ITR-flanked transgene-containing plasmid and a construct coding for Rep and Cap in adenovirus-infected cells (Hermonat and Muzyczka 1984; Tratschin et al. 1984b). These early vectors were contaminated with wild type (WT) AAV and heat-inactivated adenovirus, nevertheless they showed that it was possible to transduce cells to achieve transgene expression from foreign DNA packaged in AAV particles. Further refinements in AAV vector production (Samulski et al. 1989; Flotte et al. 1995) that allowed for the generation of vector stocks free of WT AAV and adenovirus contaminants led to the first *in vivo* gene therapy study using rAAV in 1993 (Flotte et al. 1993). This was followed few years later by the initiation of the first clinical trial in cystic fibrosis patients (Flotte et al. 2003). In the last 20 years, the study of AAV-based vectors for gene therapy has attracted a lot of interest and the field has expanded rapidly, generating a vast amount of knowledge that translated into numerous clinical trials.

In parallel to the development of AAV vectors for gene therapy studies, investigations were ongoing to characterise the integration of AAV in the host cell genome. The initial observations on latency were confirmed in several cell lines using both WT and recombinant AAV (Laughlin et al. 1986; McLaughlin et al. 1988). Remarkably, the analysis of the WT AAV integration sites revealed that most integration events mapped to a site on human chromosome 19, which was termed *AAVS1* (Kotin et al. 1990; Kotin et al. 1991; Samulski et al. 1991; Kotin et al. 1992). The determinants for site-specific integration were uncovered soon after and

included the large Rep proteins (Balague et al. 1997; Surosky et al. 1997), and sequences shared by *AAVS1* and the viral ITRs, namely the Rep binding site (RBS) and the terminal resolution site (TRS) present in the *AAVS1* site, and the viral RBS (Weitzman et al. 1994; Urcelay et al. 1995; Linden et al. 1996a). It was initially thought that both WT and recombinant AAV could integrate site-specifically (Samulski et al. 1989). However, it later became clear that in the absence of Rep, AAV vectors only persist as circular episomes (Schnepp et al. 2003) or integrate randomly at very low frequency (Walsh et al. 1992; Kearns et al. 1996). Recent developments in the study of AAV integration, facilitated by the advent of more advanced PCR-based technologies, are now shedding additional light on the mechanism of integration (reviewed in section 1.1.3.8).

Today, despite the fact that much remains to be learned about AAV, research on the basic virology of AAV is overshadowed by its promise as a gene therapy vector. The process of AAV site-specific integration and its implications, including whether site-specific integration is actually part of the AAV life cycle or a mere consequence of the biochemical activities of the Rep proteins, for instance, are starting to become more accessible as advanced DNA sequencing technologies become available. Furthermore, the remarkable multi-functionality of the Rep proteins (Mendelson et al. 1986; Im and Muzyczka 1990) and the mechanisms by which they orchestrate the AAV life cycle are still under investigation. Understanding the full extent of their versatile biochemical properties, including their ability to site-specifically bind and nick DNA substrates, may reveal an exciting biotechnological potential. Finally, the study of the interactions between AAV and its helper viruses may provide insights into virus-virus interactions, an undervalued aspect of biology that could have important implications in human health are only beginning to emerge (Virgin 2014).

1.1.2 AAV – the virus

1.1.2.1 Taxonomy

AAV belongs to the Parvovirus family, which includes all small, non-enveloped, linear single-stranded DNA viruses. The *Parvoviridae* are remarkably widespread throughout the animal kingdom, and are further classified into the subfamilies

Parvovirinae, infecting vertebrates, and *Densovirinae*, infecting arthropods (Tattersall 2006). AAV is further classified into the dependovirus genus, based on its dependence on helper functions provided by other DNA viruses, such as adenovirus or herpesvirus, to complete a productive replication cycle.

To date thirteen AAV serotypes (AAV1 to 13) have been described (Srivastava et al. 1983; Muramatsu et al. 1996; Chiorini et al. 1997; Rutledge et al. 1998; Chiorini et al. 1999b; Gao et al. 2002; Gao et al. 2004; Mori et al. 2004; Schmidt et al. 2006; Schmidt et al. 2008a; Schmidt et al. 2008b), and serological studies have detected antibodies against AAV in human and non-human primates (Blacklow et al. 1968; Chirmule et al. 1999; Kerr and Linden 2006; Calcedo et al. 2009). Furthermore, a vast number of *cap* gene sequences have been identified in primate tissues. AAV sequences have also been described for other mammals (dogs, pigs, cows and horses) and in birds. The current taxonomy of AAV is based on *cap* gene sequence similarity, as it is clear that each isolate does not represent a distinct serotype. AAV sequences have been grouped into six clades (A-F) and two clonal isolates (AAV4 and AAV5) (Gao et al. 2004). It is noteworthy that most of the knowledge on AAV biology comes from AAV2, which is the most extensively studied AAV serotype and also the focus of this thesis.

1.1.2.2 Genome structure

The AAV genome (Figure 1) is a single stranded linear DNA (ssDNA) molecule of 4.7kb; both plus and minus strands are infectious (Samulski et al. 1987) and are packaged with equal efficiency into separate preformed capsids (King et al. 2001). ITRs – imperfect palindromic sequences that self-anneal to form T-shaped hairpins (Lusby et al. 1980) – flank the AAV genome and provide all the *cis*-acting elements necessary for replication, packaging and integration (Samulski et al. 1989), including the RBS and the TRS.

The AAV genome contains two major ORFs: the left ORF encodes for the non-structural Rep proteins, while the structural Cap proteins are encoded from the right ORF (Srivastava et al. 1983) (Figure 1). An additional open reading frame, located within the VP2 coding sequence, encodes for a protein involved in capsid assembly called assembly-activating protein (AAP)(Sonntag et al. 2010). AAV

exploits a combination of different promoters (p5, p19 and p40, identified by their relative position in the viral genome), alternative splicing and non-conventional start codons to make the most of its limited coding capacity (Lusby et al. 1980; Marcus et al. 1981; Qiu et al. 2006). The p5 promoter gives rise to the large Rep proteins, Rep78 and Rep68, while the p19 promoter controls the expression of the smaller Rep52 and Rep40. All the capsid proteins, VP1, VP2 and VP3, are expressed from the p40 promoter (Figure 1).

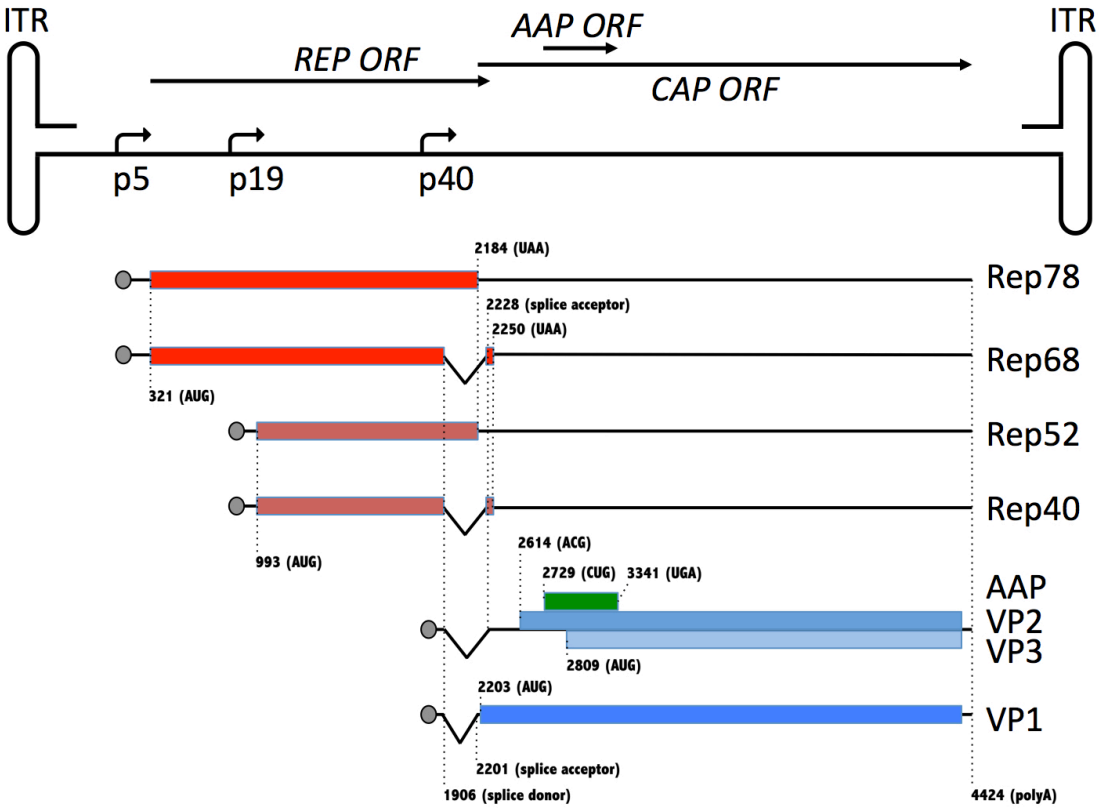


Figure 1: AAV genome, genes, transcriptional units and proteins.
The genome of AAV is a single-stranded DNA molecule that is flanked by T-shaped ITRs. It contains three ORFs: the REP ORF coding for the non-structural Rep proteins, the Cap ORF coding for the structural capsid proteins and the AAP ORF. There are three promoters, identified by their relative position in the genome: p5, p19 and p40. The p5 promoter controls the expression of the large Rep proteins Rep78 and the spliced variant Rep68. The p19 gives rise to the small Rep proteins Rep52 and Rep40. The p40 promoter controls the three capsid proteins VP1, VP2 and VP3, as well as the AAP. There is a single splice donor site and two splice acceptor sites, one of which (2228) is used most frequently. Alternative splicing and non-conventional start codons (for VP2 and AAP) allow the generation of 8 proteins (coloured boxes). All the AAV mRNAs use a single polyA tail located at the right end of the genome.

1.1.2.3 Capsid structure

The AAV capsid protects the infectious AAV genome throughout the AAV life cycle. However, the capsid is more than a simple shell, it provides important functions necessary for the correct trafficking of the virus. These functions include host cell recognition and binding, entry and trafficking to the nucleus, release of the viral genome at the appropriate time and place, egress from the host cell, and escape from immune surveillance (Agbandje-McKenna and Kleinschmidt 2011). The AAV 4.7kb genome is packaged inside a small T=1 icosahedral capsid of around 260Å (26nm) in diameter. A total of 60 copies of VP1, VP2 and VP3, in a ratio of 1:1:10, form the AAV virion. The three structural proteins have an identical C-terminal sequence, involved in most capsid functions, although VP1 and VP2 have additional functionally relevant N-terminal amino acids (AAs). The unique N-terminal VP1 residues contain a phospholipase A_2 (PLA_2) activity that is necessary for endosomal escape (Girod et al. 2002; Sonntag et al. 2006; Stahnke et al. 2011), while the shared VP1/VP2 sequence contains a nuclear localisation signal (Grieger et al. 2006b; Xiao and Samulski 2012). In addition, AAV capsids have been shown to have a pH-dependent protease activity, although its role in viral infection remains unclear to date (Salganik et al. 2012).

The capsid structures of AAV serotypes 1 to 9 have been solved (Xie et al. 2002; Walters et al. 2004; Padron et al. 2005; Miller et al. 2006; Nam et al. 2007; Quesada et al. 2007; Lerch et al. 2010; Ng et al. 2010; DiMattia et al. 2012), but only the VP3 common region has been observed. The additional VP1 and VP2 residues appear disordered, and it has been suggested that they are situated on the inside of the capsid (Kronenberg et al. 2005). Part of the VP3 structure, consisting of an eight-stranded antiparallel β-barrel core and a short α-helix is highly conserved amongst AAV serotypes. The majority of the VP3 protein sequence, however, is made of loops that connect the core β-strands, including regions that account for serotype variability (Agbandje-McKenna and Kleinschmidt 2011). These regions, called variable regions (VRs), are exposed on the outside of the capsid and are associated with the capsid functional roles, including receptor binding and tissue tropism, and also determine capsid antigenicity. This is of particular relevance for the design of

gene therapy vectors that target disease-specific tissues or cell types (Drouin and Agbandje-McKenna 2013).

The assembled icosahedral capsid structure brings together VP subunits at two-, three- and five-fold symmetry axis. The two-fold axis, formed by the interaction of two VP monomers, creates a depression in the capsid surface and has been suggested to be involved in structural changes occurring during endosomal trafficking (Nam et al. 2011). Three VP subunits strongly interact at the three-fold axis of symmetry, implicated in receptor binding and antibody recognition, to form protrusions surrounding a central depression (Agbandje-McKenna and Kleinschmidt 2011). Finally, the five-fold axis interaction creates a central channel surrounded by a depression, and is the only point of contact between the inside and the outside of the capsid. The packaging of the viral DNA and the externalisation of the VP1/VP2 enzymatic functions are thought to occur at this interface (Kronenberg et al. 2005; Bleker et al. 2006).

1.1.3 The AAV life cycle

Despite its relatively simple genomic architecture, AAV has a complex and tightly regulated life cycle, consisting of a latent phase in the absence of helper co-infection and a productive replicative phase in its presence (Figure 2). In both cases, the infectious AAV particles are trafficked to the nucleus, where the DNA is released from the capsid (see section 1.1.3.2). Under permissive conditions, such as co-infection with adenovirus or herpes simplex virus (HSV), the AAV DNA will undergo several rounds of replication (see section 1.1.3.4) before being packaged into newly formed capsids (see section 1.1.3.5). Under non-permissive conditions AAV will remain latent in the cell, either episomally or integrated in the host genome through a mechanism that requires the Rep proteins (discussed in section 1.1.3.8).

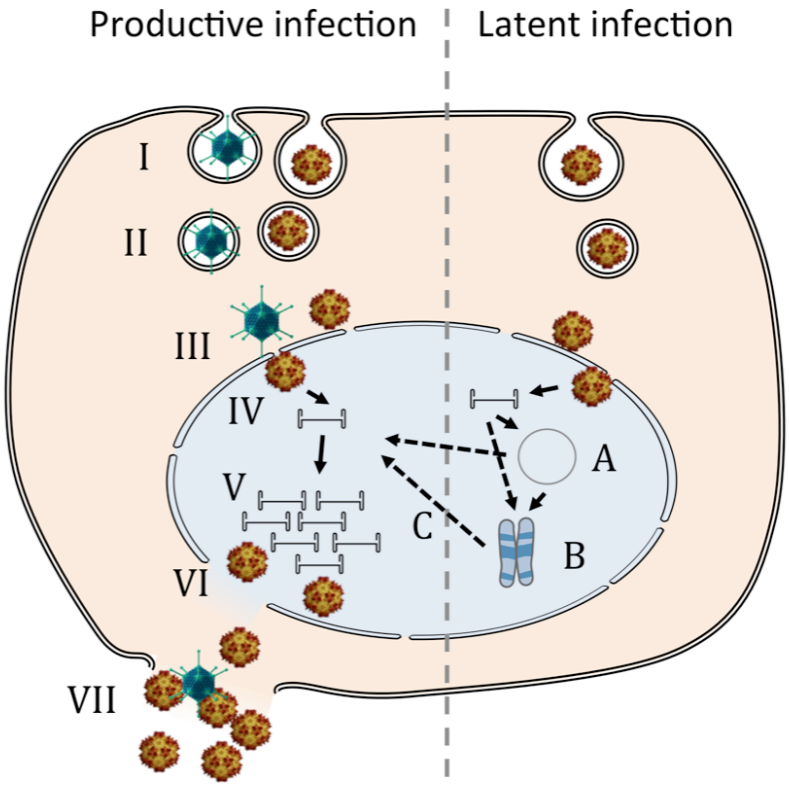


Figure 2: The AAV life cycle.
The AAV life cycle is divided between productive infection in the presence of helper virus (adenovirus in the picture) and latent infection in its absence. (I) The AAV particle enters the cell by receptor-mediated endocytosis and (II) is trafficked through the endosomal system to (III) the nuclear membrane (discussed in section 1.1.3.2). (IV) After translocation to the nucleus, the viral DNA is released and (V) undergoes several rounds of replication (section 1.1.3.4) before being (VI) packaged in pre-formed capsids and (VII) egressing from the cell (section 1.1.3.5). In the absence of helper functions, the AAV particle reaches the nucleus where it can establish latency by (A) forming a stable circular episome or (B) by integrating into the human genome (discussed in section 1.1.3.8). In the event of helper virus superinfection, (C) the AAV genome can be rescued from its latent state and undergo productive replication. The relative sizes of the elements of this diagram are not in scale but are adapted for clarity.

1.1.3.1 AAV infection in humans

Despite widespread AAV infection of the human population (Chirmule et al. 1999; Erles et al. 1999; Calcedo et al. 2009), it is widely accepted that human AAV infection is not linked to any pathology (Kerr and Linden 2006). The lack of pathogenicity has hindered the study of the viral life cycle *in vivo*, and most of our knowledge of the AAV life cycle, comes from tissue culture studies. A few studies, however, have attempted to investigate WT AAV infection in its human host. AAV has been isolated from several tissues that are common sites of infection by AAV

helper viruses, including respiratory, gastrointestinal and genital tract tissues (Blacklow et al. 1967; Bantel-Schaal and zur Hausen 1984; Friedman-Einat et al. 1997). Another study detected AAV sequences in the muscle (Tezak et al. 2000), a site where helper virus infection is not expected and thus could represent a potential reservoir for AAV latency. None of these studies, however, investigated whether the viral DNA was present as an episome or had integrated into the host genome. More recently, the group of Philip Johnson characterised the molecular structure of AAV isolated from tonsil and adenoids as well as from spleen and lung tissues (Chen et al. 2005; Schnepp et al. 2005; Schnepp et al. 2009). In these tissues, AAV genomes only persisted as circular episomes that could be isolated and were shown to be infectious upon transfection in the presence of adenovirus. Finally, epidemiological studies have attempted to correlate AAV infections with protection from certain cancers or with infertility, however contrasting conclusions have been reached (Kerr and Linden 2006).

1.1.3.2 Attachment, entry and intracellular trafficking

AAV infection begins with the attachment of the infectious viral particles to host cell surface receptors. Different AAV serotypes have been shown to use distinct cell surface glycans as primary receptors (Agbandje-McKenna and Kleinschmidt 2011; Nonnenmacher and Weber 2012), and thus can infect a variety of tissues *in vivo*. Heparan sulphate proteoglycan (HSPG) was the first AAV receptor identified (Summerford and Samulski 1998), and binds AAV2, AAV3 and AAV6. AAV2 variants based on sequences detected in humans, however, do not bind HSPG, suggesting that AAV2 binding to HSPG is a tissue culture adaptation (Chen et al. 2005), highlighting a potential limitation of tissue culture studies. A specific capsid region at the three-fold axis of symmetry seems to be the site for receptor binding, and differences in this region are responsible for cell surface glycan specificity (Agbandje-McKenna and Kleinschmidt 2011). Following primary receptor attachment, AAV particles interact with secondary co-receptors before being internalised. Co-receptors are generally proteins rather than glycans, and include integrins, fibroblast growth factor receptor, hepatocyte growth factor receptor, and laminin (Agbandje-McKenna and Kleinschmidt 2011; Nonnenmacher and Weber

2012). Furthermore, it has been suggested that following attachment, the AAV capsid undergoes conformational changes that could facilitate interaction with secondary receptors (Levy et al. 2009). The AAV particles are then internalised through endocytosis. Different internalisation pathways have been described for AAV2. While clathrin-mediated endocytosis has long been thought to be the main pathway, recent observations suggest that the clathrin-independent carriers/GPI-enriched endocytic compartment (CLIC/GEEC) pathway might also play a significant role (Nonnenmacher and Weber 2011). In addition, AAV5 can also be internalised through caveolar endocytosis (Bantel-Schaal et al. 2009).

The intracellular trafficking of AAV to the nucleus is a slow and rate-limiting process (Agbandje-McKenna and Kleinschmidt 2011; Nonnenmacher and Weber 2012). AAV particles have been shown to localise to early, late and recycling endosomes, as well as lysosomes. In addition, transport to the Golgi apparatus has also been reported. However, it is not entirely clear which pathway leads to efficient infection, and serotype differences are expected. Endosomal processing is necessary for efficient AAV transduction, and causes conformational changes to the AAV capsid, possibly inducing the externalisation of the PLA_2 domain of VP1 (Nonnenmacher and Weber 2012). Further modifications of the capsid occur after escape into the cytoplasm, including the exposure of the nuclear localisation signal present in VP1 and VP2. Inhibition of capsid phosphorylation (Zhong et al. 2008) and ubiquitination (Yan et al. 2002) has also been suggested to increase transduction efficiency, although the precise mechanisms behind this effect are not clear. The viral particles are then transported to the nucleus, where a series of poorly understood events take place. First, the virus translocates across the nuclear membrane via the nuclear pore complex, in a process mediated by the nuclear localisation signal present in VP1 and VP2 (Grieger et al. 2006b; Popa-Wagner et al. 2012) and interactions with the nuclear pore components (Nicolson and Samulski 2014). Once in the nucleus, the intact viral particle is rapidly transported to the nucleolus, from where it can be mobilised into the nucleoplasm (Johnson and Samulski 2009). Finally, the AAV genome is released by an unknown mechanism (Agbandje-McKenna and Kleinschmidt 2011; Nonnenmacher and Weber 2012).

While the understanding of AAV trafficking through the host cell has improved significantly over the last years, many important details remain to be elucidated. Several aspects of AAV biology, however, complicate these analyses. First, most studies have been carried out in the absence of helper virus, thus reflecting the initiation of a latent infection rather than a productive one. In addition, host cell infection by a helper virus is likely to change the overall cellular environment and therefore could have consequences on the trafficking of AAV. Furthermore, it is plausible that the trafficking pathway used by AAV will be dependent on the cell type being infected and the multiplicity of infection (MOI) used to infect that cell, as well as the AAV serotype. Finally, only a fraction of the AAV particles that enter a cell is infectious. This phenomenon is particularly marked for rAAV vectors (Zeltner et al. 2010), which have been used in all trafficking studies. Distinguishing between the infectious particles that will cause productive viral replication and the non-infectious ones is a significant challenge that has not been resolved.

1.1.3.3 Gene expression and transcriptional regulation

Once the DNA is released inside the nucleus, the ssDNA genome is converted to dsDNA, a prerequisite for transcription and gene expression (Ferrari et al. 1996). As described above, three viral promoters, identified by their relative position in the genome, give rise to 6 overlapping RNAs (see Figure 1) (Qiu et al. 2006). In AAV2, all these RNAs are polyadenylated at the same site at the right end of the genome, and contain a single intron that can use one of two closely located splice acceptors. Unspliced p5- and p19-derived messengers code for Rep78 and Rep52, respectively, while their spliced counterparts encode for Rep 68 and Rep40. Two spliced mRNAs generated from the p40 promoter generate the structural proteins VP1, VP2 and VP3. The mRNA generated using the preferred splice acceptor site gives rise to VP3 and, through the use of a non-conventional start codon, to VP2. The larger VP1 protein is translated from the minor spliced message.

Despite the fact that AAV RNAs are stable, their relative abundance and the ratio of spliced to unspliced RNAs, as well as their temporal order of appearance during an infection, are tightly regulated (Mouw and Pintel 2000). During a productive infection, i.e. during co-infection with a helper virus, p5-generated RNAs are the

first to appear, followed by p19 transcripts, ensuring efficient replication of the AAV DNA. The p19-generated RNAs accumulate to higher levels than those from p5, and later during the infection the p40 transcripts form the majority of AAV RNA species, skewing the equilibrium towards the production of new infectious particles rather than replication. In addition, the ratio of spliced to unspliced RNAs also increases during infection.

In the absence of helper virus co-infection, expression of the AAV genes is very low (Laughlin et al. 1986; Trempe and Carter 1988), and the viral promoters are repressed by the combined action of the Rep proteins and cellular factors such as YY1 and MLTF (Labow et al. 1986; Beaton et al. 1989; Chang et al. 1989; Shi et al. 1991; Weger et al. 1997). The levels of p5-generated transcripts are optimal to support Rep-mediated integration into the host genomic DNA, but sufficiently low to prevent Rep-mediated rescue of the integrated DNA. The Rep-mediated repression of the p5 promoter has been shown to require both the RBS in p5 (McCarty et al. 1994a) and the ATPase activity of Rep (Kyostio et al. 1995), while repression of the p19 promoter only depends on the ATPase activity (Kyostio et al. 1994; Horer et al. 1995). Upon helper virus co-infection, the repression of the p5 promoter by the cellular transcription factors YY1 and MLTF is lifted (Chang et al. 1989; Shi et al. 1991), leading to a controlled transactivation of all the viral promoters to achieve the correct protein expression (Labow et al. 1986; McCarty et al. 1991; Pereira and Muzyczka 1997a; Pereira and Muzyczka 1997b; Weger et al. 1997; Lackner and Muzyczka 2002). The Rep proteins have been shown to act both as transcriptional activators and repressors (Labow et al. 1986; Pereira et al. 1997), whereby a feedback loop has been suggested to maintain a precise ratio of p19 to p5 transcripts. More specifically, Rep binding to either the ITR or the p5 RBS mediates activation of p19 and p40 transcription, while expression from the p5 promoter is enhanced through Rep binding to the ITR RBS and repressed by the p5 RBS. The small Rep proteins, Rep40 and Rep52, can, in turn, lift repression from the p5 promoter. In addition, the large Rep proteins have also been shown to increase the ratio of spliced to unspliced RNAs through binding to an extended RBS, provided either by the p5 promoter or the ITR (Qiu and Pintel 2002).

1.1.3.4 DNA replication

The basic model for parvovirus and AAV replication was described several decades ago (Straus et al. 1976; Tattersall and Ward 1976), and it is still accepted today (Ward 2006). The mode of replication is often referred to as rolling hairpin replication, a modified version of plasmid rolling circle replication, and implies a strand displacement mechanism (Figure 3).

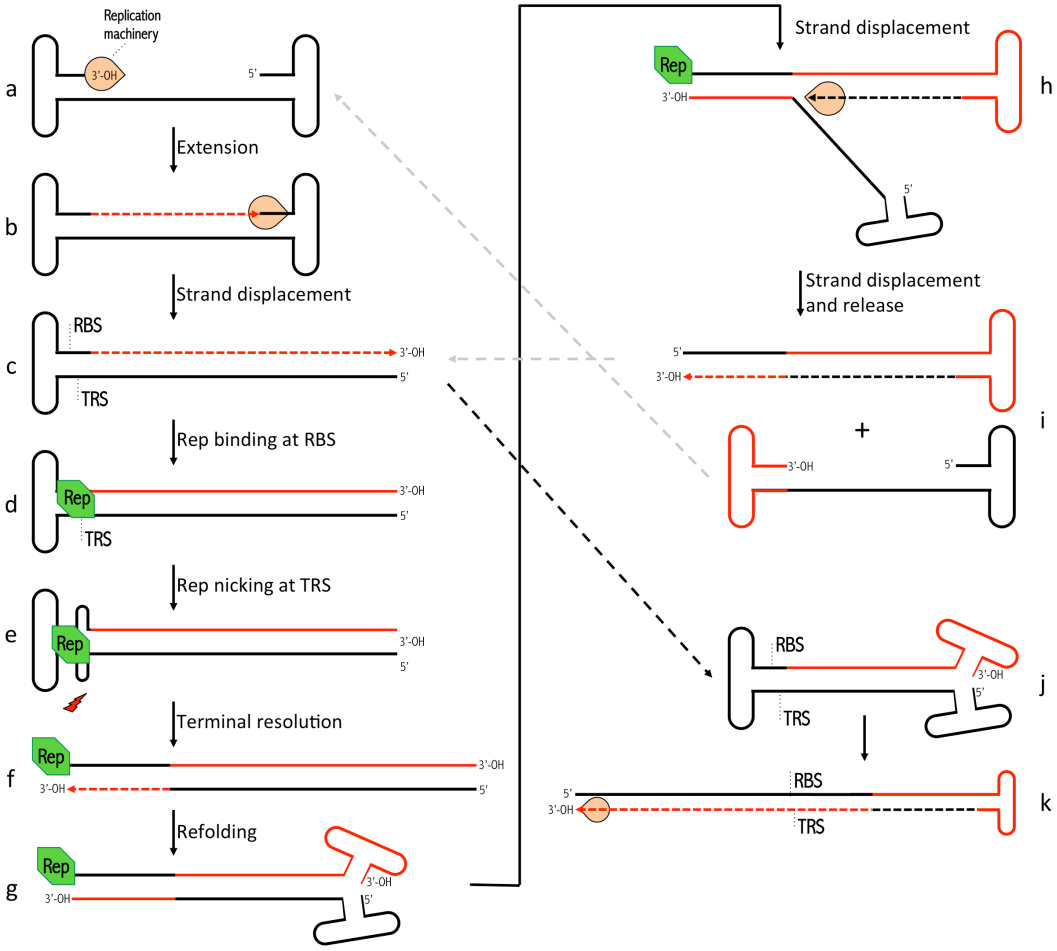


Figure 3: Rolling hairpin replication of AAV DNA.
The free 3'-OH provided by the self-annealed ITR (a) serves as the primer for DNA replication mediated by the cellular machinery (b). The right ITR is replicated by strand displacement, generating a ds molecule covalently linked by the left ITR (c). Rep binds at the RBS in the left ITR (d) and nicks at the nearby TRS site (e) to allow the terminal resolution of the left ITR (f). The complementary ITR sequences at the right end of the genome fold back (g), freeing a new 3'-OH that can be used for another round of replication by strand displacement (h) that generates a new ss AAV genome and a ds molecule covalently linked by the right ITR (i). These molecules can then be used for another round of replication (grey arrows) or potentially could serve as packaging templates. If the refolding of the right ITR occurs prior to the terminal resolution of the left ITR (j) followed by replication by strand displacement, a dimeric replication intermediate that is detectable in replication assays can be formed (k). By subsequent RBS binding and TRS nicking the dimeric intermediate can re-enter the replication cycle.

The first step in replication consists of the second-strand synthesis: the self-annealed ITR on the 3' end of the genome serves as a primer for unidirectional DNA synthesis (Figure 3, a and b), resulting in a double-stranded molecule covalently linked at one side (Figure 3c). The replication of the original ITR, or terminal resolution, requires binding to the RBS (Figure 3d) and nicking at the TRS (Figure 3e) by the large Rep proteins generating a new free base-paired 3'-OH group for DNA synthesis through the ITR (Figure 3f). This results in a double-stranded linear DNA molecule, the monomeric replicative intermediate. Both ends can then fold back to form double hairpinned structures (Figure 3g), and another round of replication by strand displacement can start (Figure 3h). At the end of each cycle of replication, two DNA molecules are generated: a double stranded one with one end covalently closed (as in Figure 3c), and a progeny single strand molecule (Figure 3i). In addition, if completion of refolding and strand displacement at one end occurs before the terminal resolution at the other end (Figure 3, j and k), a dimeric intermediate formed by two genomes in a head-to-head or tail-to-tail orientation can be generated (Figure 3k). This mode of replication leads to an inversion of part of the ITR sequence during each round of replication, generating two alternative conformations designed as flip and flop (Lusby et al. 1980). It is still unknown which of the possible DNA molecules constitutes the template for packaging.

The essential elements required for the efficient replication of AAV DNA include the viral origin of replication (the ITRs) (Hauswirth and Berns 1977; Lusby et al. 1980), the large Rep proteins (Rep78 or Rep68) (Im and Muzyczka 1990) and factors provided by helper-virus infected cell extracts (Hong et al. 1992; Ni et al. 1994; Ward and Berns 1996). Rep binds to the RBS present in the ITR, and makes further contact with a sequence in the hairpin termed RBS' (Snyder et al. 1993; Chiorini et al. 1994b; McCarty et al. 1994b; Ryan et al. 1996). Following binding, Rep unwinds the ITR stem in an ATP-dependent reaction to expose a single-stranded TRS for nicking (Brister and Muzyczka 2000; Smith and Kotin 2000). In addition, Rep could also provide the helicase activity necessary to unwind the ITR to allow its replication after terminal resolution (Zhou et al. 1999; Ward et al. 2001). The structure of the ITR, rather than its exact sequence, appears to be particularly important for terminal resolution (Lefebvre et al. 1984; Bohenzky et al. 1988).

Indeed, mutations of the ITR that maintain the overall hairpin structure, but that do not affect the RBS' sequence, are still replicated.

Plasmids that contain the full AAV genome are infectious; when transfected into cells that are subsequently infected with adenovirus, infectious AAV is produced (Samulski et al. 1982). The replication products obtained from these plasmids are indistinguishable from those observed after AAV infection. The viral sequences are rescued from the plasmid DNA by a Rep-mediated mechanism involving nicking at the ITR and replication by strand displacement (Samulski et al. 1982). A similar mechanism is thought to provide the basis for the release of the viral DNA integrated into the host cell genome through Rep-mediated integration after adenovirus infection (discussed in section 1.1.3.8). Deletions at one end of the genome (up to 113 bases) are repaired; clones containing such deletions are still infectious and the replication products contain full-length hairpins in both sides (Samulski et al. 1983). Simultaneous deletions in both ITRs, on the other hand, cannot be repaired. Interestingly, a plasmid containing shorter deletions of 55 nucleotides on both ends of the AAV sequences does not support rescue of the viral genome, but the presence of the RBS and TRS is sufficient to initiate replication of the entire plasmid, indicating that RBS and TRS act as minimal origins of replication (Hong et al. 1992).

In vitro, complete AAV replication can be achieved using the host proteins DNA polymerase delta (POLD), the proliferating nuclear antigen (PCNA), the replicating factor C (RFC) and the minichromosome maintenance complex (MCM) in combination with the large Rep proteins (Nash et al. 2008). In cells, and in the presence of helper virus, the situation is necessarily more complex. The replication of AAV, similarly to many DNA viruses, takes place in specialised replication centres in the nucleus of infected cells, where AAV and helper virus proteins interact with several cellular proteins to efficiently replicate the AAV DNA (Vogel et al. 2013). The composition of the AAV replication centres is not fixed during the life cycle: it is likely to change spatially and temporally over the course of replication and also depends on helper virus functions. Both adenoviruses and herpes viruses are efficient helpers of AAV replication (discussed in more detail in section 1.1.3.6). Other viruses, such as vaccinia virus (Schlehofer et al. 1986) and papilloma virus

(Walz et al. 1997) have also been shown to be able to support AAV replication, suggesting that the helper viruses alter the cellular environment and make the cell permissive for AAV replication, rather than providing specific replication functions. Accordingly, cell treatment with certain genotoxic agents can render cells permissive for sub-optimal AAV replication in the absence of helper viruses (Yakobson et al. 1987; Yalkinoglu et al. 1988; Yakobson et al. 1989).

1.1.3.5 Packaging and egress from the cell

The VP subunits that form the AAV capsid assemble into empty capsids in the nucleoli of infected cells (Wistuba et al. 1997), where they interact with the AAP (Sonntag et al. 2010). The progeny ssDNA is then slowly translocated into the preformed capsid (Myers and Carter 1980). The ITRs act as minimal *cis* elements (McLaughlin et al. 1988), although the precise packaging signal has not yet been identified. The small Rep proteins, and in particular their helicase function, are required for the efficient encapsidation of viral DNA and the accumulation of ssDNA, which correlates with an increase in DNA-containing particles (Chejanovsky and Carter 1989; King et al. 2001).

The egress of AAV particles from cells is very poorly characterised. Like most non-enveloped viruses, AAV is thought to be released from infected cells after accumulation of viral particles and lysis of nuclear and cellular membranes. However, infectious AAV particles can be detected in the tissue culture medium before the appearance of a cytopathogenic effect (CPE) in AAV-producing cells, although the degree of secretion varies amongst serotypes (Vandenberghe et al. 2010). In addition, active nuclear exit of viral capsids and outward transport associated with virion maturation has been demonstrated for another parvovirus, the minute virus of mice (MVM) (Valle et al. 2006; Bar et al. 2013), suggesting the possible existence of an alternative release pathway that may be functionally relevant for AAV as well.

1.1.3.6 Helper functions

Only a subset of helper virus proteins is necessary for complete support of the AAV life cycle; the best characterised are the helper functions of adenovirus (Ad) and herpes simplex virus 1 (HSV-1) (Geoffroy and Salvetti 2005; Vogel et al. 2013).

In the case of Ad, the helper factors necessary to support AAV replication are the viral proteins E1A, E1b55K, E2A, E4orf6, and the viral associated RNA (VA RNA). The E1A protein activates the transcription of adenoviral early genes and, similarly, induces transcription of the AAV Rep ORF (Chang et al. 1989). The E1b55K and the E4orf6 proteins form a complex that plays a role in the nuclear export and in the metabolism of AAV mRNA (Mouw and Pintel 2000), as well as in second-strand synthesis (Ferrari et al. 1996). In addition, the E1b55K/E4orf6 enhances AAV replication by degrading the Mre11 complex (MRN) (Schwartz et al. 2007). The MRN complex, composed of Mre11, Rad50, and Nbs1 is a conserved multi-protein complex that is involved in DNA damage sensing and repair (D'Amours and Jackson 2002). E2A encodes for the DNA-binding protein (DBP) and supports multiple steps of the viral life cycle, including transcriptional regulation, mRNA processing, and replication (Janik et al. 1989; Ward et al. 1998). Finally, the VA RNAs have been shown to maintain AAV protein translation (Janik et al. 1989).

The helper functions encoded by HSV-1, unlike those of Ad, include proteins that are also necessary for HSV-1 replication. These include the helicase-primase complex encoded by UL5, UL8 and UL52, and the DNA binding protein ICP8, product of the UL29 gene (Weindler and Heilbronn 1991). The role of UL52 was later revised, as it was shown that its primase activity is not required for AAV replication (Slanina et al. 2006). Further investigations also revealed that the ICP8 protein plays a similar role to the adenoviral DBP in enhancing AAV replication, and additionally localises to AAV replication centres and can bind Rep78 *in vitro* (Heilbronn et al. 2003; Stracker et al. 2004). The HSV-1 DNA polymerase complex formed by the polymerase UL30 and the dsDNA binding protein UL42, has been shown to enhance AAV replication (Alazard-Dany et al. 2009). Furthermore, the HSV-1 polymerase UL30 can replicate the AAV genome efficiently *in vitro*, replacing the cellular replication machinery (Ward et al. 2001). HSV-1 proteins involved in the transcriptional regulation of the AAV proteins have also been identified (Alazard-Dany et al. 2009). Finally, a recent report described how MRN degradation, which is beneficial to AAV replication in the presence of adenovirus, is detrimental when AAV replication is supported by HSV-1 (Millet et al. 2015).

1.1.3.7 AAV interactions with cellular factors

Several cellular proteins have been described that interact with the Rep proteins, with the AAV genome, or more generally involved in the AAV life cycle (Vogel et al. 2013). The exact contributions of many of these host proteins to the AAV life cycle, however, are not known. Furthermore, the range of host proteins involved is different depending on which helper functions are present. Proteins identified include factors involved in DNA replication, transcriptional regulation, DNA repair and DNA damage response (DDR), and RNA metabolism. In addition, cytoplasmic and mitochondrial proteins have also been found to interact with Rep.

Some of the best-characterised host factors required for the AAV life cycle are involved in DNA replication. *In vitro*, POLD, PCNA, RFC, and MCM, together with Rep, are sufficient to replicate AAV DNA (Nash et al. 2008). In the presence of HSV-1 co-infection, it is thought that the HSV-1 polymerase can substitute POLD. DNA binding proteins, either cellular (RPA) or viral (DBP for Ad and ICP8 for HSV-1), further enhance the *in vitro* replication (Ward et al. 1998; Stracker et al. 2004). The high mobility group protein 1 (HMG1) has been suggested to support Rep activities, and thus is likely to be involved in AAV replication (Costello et al. 1997). On the other hand, FKBP52, a protein that binds the D-region within the viral ITRs, has been shown to inhibit replication by blocking second strand synthesis in the absence of helper virus (Qing et al. 2001). In addition, the AAV genome, and likely its replication intermediates, are recognised by proteins of the DNA repair and DDR machineries (Vogel et al. 2013). In particular, components of the MRN complex and of the ATM/DNA-PK signalling pathways have both been shown to be recruited to AAV replication centres. However, whether these pathways of DNA damage sensing and repair are beneficial or detrimental to the viral replication is not always trivial, and may depend on the helper virus present. For example, the knock-down of MRN has been shown to have opposite effects in the presence of adenovirus or HSV-1 (Schwartz et al. 2007; Millet et al. 2015). In addition, DDR proteins have also been shown to interact with recombinant AAV vectors in the absence of helper viruses (Choi et al. 2006; Cervelli et al. 2008).

1.1.3.8 Latent infection and Rep-mediated integration

In the absence of helper virus, the expression of Rep proteins is minimal and the AAV genome is not replicated. Nevertheless, long-term persistence of AAV genomes has been widely described (Schultz and Chamberlain 2008). Two main mechanisms for this long-lasting persistence have been identified: Rep-mediated site-specific integration and episomal persistence (McCarty et al. 2004). Rep-mediated integration, however, is difficult to study *in vivo* and has only been observed in tissue culture. Thus, the significance of this phenomenon in the AAV life cycle remains unclear.

In the early 1990's, the identification of viral-cellular junctions in cell lines latently infected with AAV led to the surprising observation that the cellular sequences all mapped to one region on the long arm of human chromosome 19 (Kotin et al. 1990; Kotin et al. 1991; Samulski et al. 1991). This target locus, subsequently termed *AAVS1*, contains an actively transcribed and ubiquitously expressed gene that encodes for the protein phosphatase I regulatory inhibitor subunit 12C (PPP1R12C), also know as myosin-binding subunit 85 (MBS85) (Dutheil et al. 2000; Lamartina et al. 2000b). MBS85, like other MBS proteins, participates in the regulation of the myosin light chain phosphatase that governs the actin-myosin assembly and disassembly. The *MBS85* gene is closely linked to three genes, *TNNI3*, *TNNT1* and *EPS8L1*, which are also involved in the regulation of actin-myosin interactions (Dutheil et al. 2000). Interestingly, orthologues of the human *AAVS1* locus have been identified in non-human primates and in rodents (Samulski et al. 1991; Dutheil et al. 2004).

Integration into *AAVS1* is not observed when cells are transduced with viral vectors devoid of Rep (Walsh et al. 1992), but can be restored when the large Rep proteins are provided *in trans* (Balague et al. 1997; Surosky et al. 1997). Further studies of *AAVS1* reveal that it contains functional homologues of the viral RBS (Weitzman et al. 1994) and TRS motifs (Urcelay et al. 1995), located immediately upstream of the *MBS85* translation start codon; these cellular motifs are necessary and sufficient to mediate integration (Giraud et al. 1994; Linden et al. 1996a). The only *cis*-acting viral sequence necessary for integration is the viral RBS, which can be provided by the ITRs or the p5 promoter (Young and Samulski 2001; Philpott et

al. 2002; Guilbaud et al. 2008). Despite several studies characterising AAV integrants at a molecular level, the definition of a consensus proviral structure has been hindered by the absence of a standardised experimental strategy (Dutheil and Linden 2006). Most of the integrants described to date are proximal to the TRS and RBS motifs, although integrated AAV sequences have also been found at a significant distance. In addition, while bi-allelic integration has not been observed, extensive rearrangements of the MBS85 gene, as well as deletions of parts of the viral sequences, are common (Dutheil and Linden 2006). Importantly, a more recent study observed that the integration event causes the partial duplication of the MBS85 sequences, and that the duplicated sequences are maintained in the 5'-3' orientation (Henckaerts et al. 2009). Together, these observations led to the proposition of a model for AAV integration that parallels that of AAV DNA replication (Figure 4) (Linden et al. 1996b; Henckaerts et al. 2009). In a first step, simultaneous binding of the cellular and viral RBS by Rep brings the AAV donor DNA in close proximity to *AAVS1* (Weitzman et al. 1994). Subsequently, the Rep protein (bound to the cellular RBS) nicks the target cellular TRS, creating a free 3'-OH group that serves as a primer for DNA replication by the cellular replication machinery. The next step implies a template switch to a nearby AAV genome, creating a junction with the left ITR, followed by one or more rounds of replication of the AAV genome. The newly replicated viral DNA is then presumably ligated to the displaced strand. Finally, the integration can be completed by the introduction of a second nick in the template strand followed by replication of the newly generated cellular-viral sequences, resulting in a partial duplication of the *AAVS1* sequence.

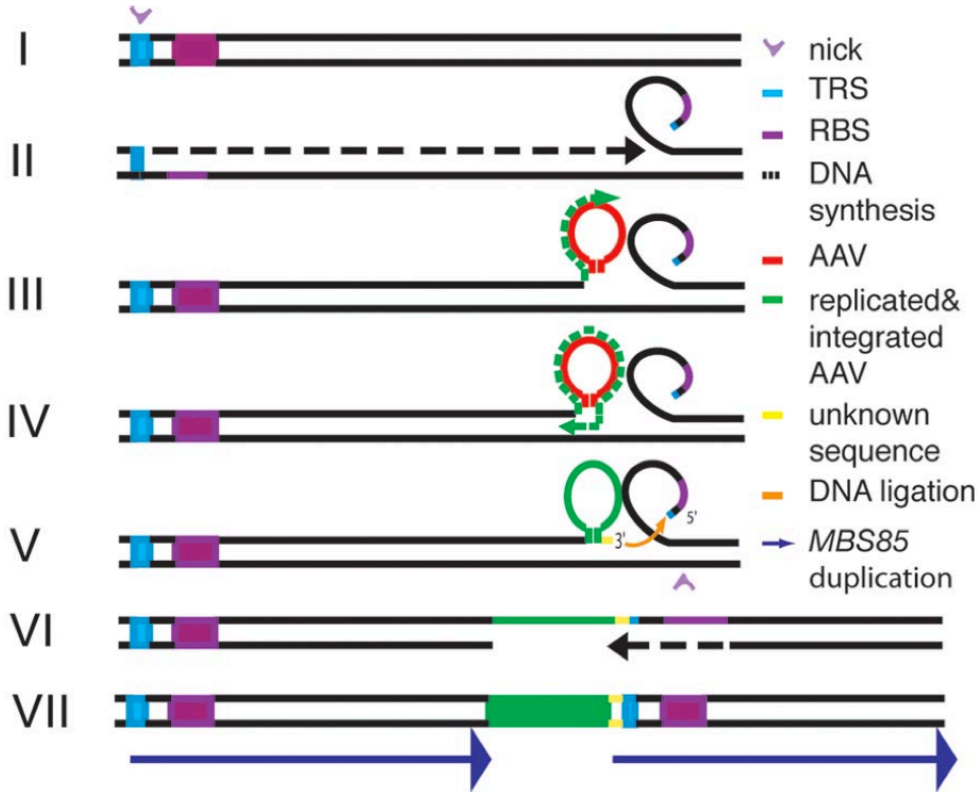


Figure 4: Model for Rep-mediated integration.
(I) Rep binds to the cellular RBS and inserts a strand-specific nick at the TRS in AAVS1. (II) DNA synthesis by strand displacement is initiated at the nicked TRS. (III) The replication switches template onto AAV. (IV) Occasional second template strand switch-back onto AAVS1 generate an inverted repeat. (V) Ligation between AAV and the displaced strand. (V and VI) Rep-mediated nicking at the bottom strand stimulates repair of the non-complementary strand. (VII) AAV site-specific integration results in partial duplication of MBS85 sequences. Adapted from (Henckaerts et al. 2009).

In recent years, the development of high-throughput sequencing technologies combined with improved PCR techniques, has allowed unbiased genome-wide studies of viral integration patterns (Schmidt et al. 2007; Paruzynski et al. 2010), including those of AAV. On one side, the results of these studies confirmed the random pattern of rAAV integration, but on the other side they challenged the notion of site-specific integration of WT AAV (Huser et al. 2010; Janovitz et al. 2013; Huser et al. 2014; Petri et al. 2015). In the first of these studies, Huser and colleagues reported that less than 10% of WT AAV integration sites detected in HeLa cells were in *AAVS1* or its proximity (Huser et al. 2010). Two other studies investigated AAV integration in HeLa cells: one reported that up to 45% of integration sites were in AAVS1 (Janovitz et al. 2013), while the other reported a value of 13% (Petri et al. 2015). In addition, two very recent reports assessed the

WT AAV integration profile in diploid human fibroblasts, and found that 2.5% (Huser et al. 2014) and 9% (Petri et al. 2015) of the integration sites were within *AAVS1*. In spite of significant differences in the experimental procedures, a general picture has emerged from these studies. First, it is clear that WT AAV integration is not as site-specific as initially hypothesised. Second, most of the integration hotspots that were analysed revealed the presence of RBS-like and/or TRS-like sequences, although *AAVS1* was the only hotspot that was consistently identified in all the studies. The putative RBS of some of the most frequent hotspots were shown to bind Rep68 *in vitro* (Huser et al. 2010; Petri et al. 2015) and, remarkably, Rep was also shown to be able to nick supercoiled DNA plasmids containing the RBS-TRS sequences from some of the most frequent hotspots (Petri et al. 2015). Thus, these data suggest that integration in sites other than *AAVS1* could be mediated by the same Rep-dependent mechanism responsible for integration into *AAVS1*. The frequency of integration in sites containing an RBS and an appropriately spaced TRS could be determined by cellular rather than viral factors. In particular, it was described that an open chromatin status correlates with an increased frequency of integration (Huser et al. 2014).

The episomal mode of AAV persistence has been most extensively investigated in the context of recombinant AAV vectors devoid of Rep, both *in vitro* and *in vivo* (McCarty et al. 2004; Schultz and Chamberlain 2008). Following the second-strand synthesis by the cellular repair machinery and/or the annealing of complementary AAV + and – strands, linear dsDNA molecules circularise and can undergo concatamerisation. In this form, episomal vector genomes have been shown to persist long-term in several tissues. Very low frequency integration of AAV vectors has also been reported (McCarty et al. 2004; Schultz and Chamberlain 2008) and it has been suggested that non-homologous recombination at double-strand breaks is the primary mechanism of this unspecific integration (Miller et al. 2004).

1.1.4 AAV as a vector for gene therapy

In recent years, AAV has become one of the most promising vectors being developed for gene therapy applications. In particular, the long-term (episomal)

persistence *in vivo* and the lack of pathogenicity associated with WT AAV infections have made it an attractive candidate as gene therapy vector.

Recombinant AAV vectors, evolved from the initial vectors developed in the mid 1980s (Hermonat and Muzyczka 1984; Tratschin et al. 1984b), are devoid of the Rep and Cap genes and are therefore replication deficient. The typical AAV vector consists of a transgene expression cassette (of maximum 4.5kb) flanked by the viral ITRs, which are the only *cis*-acting elements necessary for replication and packaging (Carter 2004). The viral Rep and Cap proteins, as well as the helper virus functions necessary for the efficient viral replication and packaging are provided *in trans*. The most common and efficient AAV vector production protocol is based on triple transfection of 293T cells with a vector plasmid, a Rep-Cap plasmid and an Ad helper plasmid (Wright 2009). While this strategy allows for efficient production of rAAVs, the infectivity of the viral vectors produced is still significantly lower than that observed for WT AAV (Zeltner et al. 2010). The need for a good manufacturing practice (GMP)-compatible and scalable production method has led to the development of alternative production platforms, such as HEK293 cell lines that grow in suspension in serum-free medium (Park et al. 2006; Durocher et al. 2007; Hildinger et al. 2007), the baculovirus expression system in insect cells (Virag et al. 2009), adeno-AAV hybrids (Zhang et al. 2009), and the use of a recombinant herpesvirus system (Clement et al. 2009). The most common strategies used to purify AAV particles are column chromatography and ultracentrifugation through non-ionic iodixanol gradients (Grieger et al. 2006a).

Over the years, a number of advances in the understanding of the biology of the virus and of AAV vectors have translated into important developments for the production of rAAVs. Beyond the discovery of the *cis*- and *trans*-acting elements of the AAV genome (Hermonat et al. 1984; Tratschin et al. 1984a), the characterisation of the minimal adenovirus helper functions allowed the development of helper plasmids carrying the adenovirus functions, abolishing the usage of WT Ad in the AAV production protocols and resulting in Ad-free rAAV stocks (Grimm et al. 1998; Matsushita et al. 1998; Xiao et al. 1998). Another significant advance was the understanding that limited Rep protein expression is beneficial to the virus and allows for increased rAAV production (Li et al. 1997). One

of the most significant advances was the development of a cross-packaging system that enables the packaging of AAV2 vector genomes (i.e. containing the ITRs from serotype 2) into other AAV capsids (Rabinowitz et al. 2002), allowing the direct comparison of the *in vivo* tropism of different AAV capsids (Rabinowitz et al. 2002; Zincarelli et al. 2008). This work was particularly relevant, as one of the ideal features of a gene-therapy vector is cell-type or tissue specificity in gene delivery and different AAV capsids have been shown to display different tissue tropisms *in vivo*. Moreover, this made possible the selection of natural or rationally designed AAV capsids with specific characteristics and tissue specificities, enabling the generation of libraries for directed evolution of capsids with the desired properties (Vandenberghe et al. 2009). Another remarkable advance was the development of double-stranded self-complementary AAVs (scAAVs) that bypass the need for second-strand synthesis once uncoated, thus allowing for a faster onset of transgene expression (McCarty 2008). While this is a very desirable property, scAAVs packaged as dimer genomes can only accommodate half of the already limited size of AAV vector genomes, and thus have a limited field of application. Finally, efforts towards increasing the capacity of rAAVs have led to the development of dual vectors that use two separate AAVs to achieve transgene expression (Ghosh and Duan 2007).

The desirable properties of rAAVs and the above-mentioned advances in vector design and production have made AAV a suitable candidate for gene therapy. In recent years, following numerous pre-clinical studies in animal models ranging from mice to non-human primates, rAAVs have entered into several clinical trials (Mueller and Flotte 2008; High and Aubourg 2011; Grieger and Samulski 2012). The initial trials investigated the use of rAAVs to treat monogenic disorders such as cystic fibrosis (Flotte et al. 2003) and haemophilia B (Kay et al. 2000), and showed very good safety profiles. However, in these initial trials, the expression of the transgene was limited and transient, most likely due to an immune response to the AAV2 capsid. Subsequent trials have targeted more immuno-privileged sites, such as the central nervous system or the retina. More recently, thanks to the cross-packaging technology described above, clinical studies are moving towards using different capsid serotypes to improve tissue targeting and to reduce the immune

responses. Today, one rAAV product has reached the market in Europe (Yla-Herttuala 2012), and several ongoing clinical trials are exploring the use of rAAVs for the treatment of a growing number of diseases.

1.2 AAV Rep and viral multifunctional proteins

1.2.1 AAV Rep proteins

1.2.1.1 The Rep proteins and their role during the AAV life cycle

The genome of AAV has evolved a single open reading frame coding for the master regulatory proteins, the Rep proteins, which are responsible for orchestrating every step in the viral life cycle. There are four Rep isoforms (Figure 5), named after their molecular weight on denaturing acrylamide gels: Rep78, Rep68, Rep52 and Rep40. As explained in section 1.1.3.3, they are generated through the usage of the p5 and p19 viral promoters and alternative splicing. The Rep proteins, in particular the large Rep proteins Rep68 and Rep78, are versatile multi-domain proteins, and are a good example of viral multi-functional proteins that have evolved under space constraints to combine several enzymatic functions within a single polypeptide chain to carry out complex and diverse biological functions.

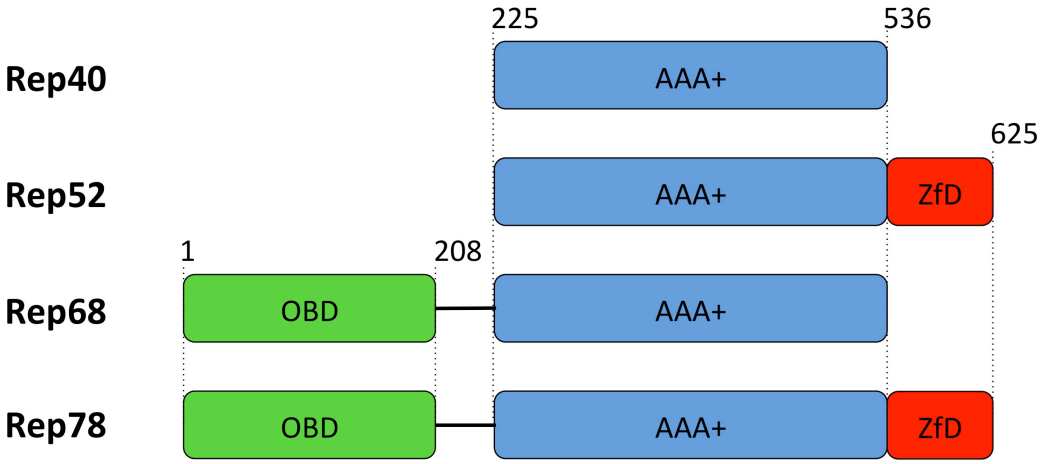


Figure 5: The Rep proteins of AAV.
Schematic diagram showing the domain architecture of the four Rep proteins of AAV. The OBD, containing DNA binding and endonuclease activities is shown in green; the central helicase domain characteristic of SF3 helicases is shown in blue, and the zinc-finger domain is shown in red.

The Rep proteins are important for the transcriptional regulation of AAV, helper virus and cellular genes, ensuring that the cellular environment is prepared for efficient replication of the viral genome or for the establishment of latency (McCarty et al. 1991; Hermonat 1994; Horer et al. 1995; Kyostio et al. 1995; Pereira et al. 1997; Weger et al. 1997; Needham et al. 2006; Dutheil et al. 2009; Dutheil et al. 2014). Transcriptional regulation can be achieved by direct binding to the RBS or other regulatory DNA sequences, and possibly through interactions with other proteins. The large Rep proteins Rep78/68 are essential for the replication of the AAV genome: their interaction with the RBS and the TRS is the basis of the terminal resolution of the ITRs and the regeneration of the 3'-OH group for subsequent rounds of replication (Ward 2006). Furthermore, Rep78/68 are required for site-specific integration of the viral genome into *AAVS1* and the rescue of the viral genome from *AAVS1* following helper virus superinfection (Dutheil and Linden 2006). The small Rep proteins Rep52 and Rep40, in addition to their role in transcriptional regulation, are important for the accumulation of single stranded genomes and their efficient packaging into preformed capsids (King et al. 2001). Finally, besides the well-characterised functional roles described above, the Rep proteins are likely to have additional and, as yet, undiscovered functions in the AAV life cycle. Rep interactions with cellular proteins, for example, remain poorly characterised, and are likely to play significant roles in the viral life cycle, including some that could improve the efficacy of AAV vectors for gene therapy.

1.2.1.2 Rep domains and their enzymatic properties

Three major domains are combined to form the largest of the Rep proteins, Rep78: an amino-terminal origin binding domain (OBD), a central helicase domain and a carboxy-terminal Zn-finger domain. A short interdomain linker (~20 amino acids) of unknown function connects the OBD and helicase domains. All the other Rep proteins are a combination of these functional domains: Rep68 is identical to Rep78 but lacks the Zn-finger domain; Rep52 combines helicase and Zn-finger domains; and finally, Rep40 only contains the helicase (Figure 5). Interestingly, this core domain that is present in all four Rep isoforms is the most conserved region amongst parvoviral non-structural proteins (Astell et al. 1987).

1.2.1.3 Origin binding domain

The amino-terminal domain of Rep, encompassing residues 1-208 and shared by Rep68 and Rep78, is termed OBD and carries sequence-specific DNA binding activity as well as strand- and sequence-specific endonuclease activity (Im and Muzyczka 1990; Owens et al. 1993; Chiorini et al. 1994a; McCarty et al. 1994a). In addition, it has also been shown to have a ligase function (Smith and Kotin 2000). The structure of AAV serotype 5 Rep OBD (residues 1-197) has been determined with and without RBS-containing DNA (Hickman et al. 2002; Hickman et al. 2004). The structural data reveal how, despite its limited size (~20kD), the OBD interacts with several DNA substrates, including the RBS, the RBS' and the TRS.

The OBD is responsible for the specific binding of Rep to RBS-containing DNA substrates, including the viral ITRs and p5 promoter, as well as the cellular integration site *AAVS1* (McCarty et al. 1994a; Weitzman et al. 1994). The RBS consists of multiple direct 5'-GCTC-3' repeats that are recognised through both major and minor groove interactions with a specialised interface of the OBD (Hickman et al. 2004). Each OBD can make contact with two adjacent repeats, and each repeat can be recognised by two separate OBD molecules on opposite sides of the DNA, forming a spiral around the RBS. The RBS sequences of different AAV serotypes and of *AAVS1* are related but not identical. They all contain at least two perfect GCTC repeats and additional imperfect repeats, yet the relative position of the base substitutions in the imperfect repeats varies. The viral p5 promoter also contains a RBS that is efficiently bound by Rep, but it only contains one perfect GCTC repeat and several imperfect repeats (McCarty et al. 1994a). In addition, *in silico* analyses of the human genome have revealed several RBS homologues, some of which have been shown to have the potential to be bound by Rep (Wonderling and Owens 1997; Lamartina et al. 2000a; Huser et al. 2010; Petri et al. 2015). Thus, the OBD-RBS interaction is specific but tolerates some degree of flexibility, which may be important for some of the Rep functions, such as the transcriptional regulation of different promoters (Batchu et al. 1994; Horer et al. 1995). In the context of the viral ITR, the OBD makes further contacts with the RBS', defined by the 5'-CTTTG-3' motif at the tip of one of the internal palindromes (Ryan et al. 1996; Hickman et al. 2004). Importantly, the position of the RBS' relative to the TRS

is maintained in both the flip and flop orientations of the internal palindromes of the ITR. During replication and site-specific integration, RBS binding is followed by nicking at the TRS. Importantly, the binding positions Rep asymmetrically on the ITR, and the correct alignment is necessary for the subsequent nicking reaction (Brister and Muzyczka 2000). Furthermore, the spacing between the RBS and the TRS also appears to be of pivotal importance for the efficient nicking at the TRS (Brister and Muzyczka 2000; Lamartina et al. 2000a). Binding to the RBS', while not essential for the nicking reaction, has been shown to increase its efficiency (Ryan et al. 1996; Brister and Muzyczka 2000).

Structurally, the AAV Rep OBD belongs to the HUH family of endonucleases, described in further detail below. The core of the AAV OBD consists of a 5-stranded anti-parallel β sheet that contains the canonical HUH motif, flanked on each side by three helices (Hickman et al. 2002) (Figure 6). Facing the two histidine residues of the HUH motif and forming the endonuclease enzymatic pocket, is helix αD, which contains the catalytically active tyrosine (Y156 in AAV2 Rep). The surface involved in RBS' binding is opposite the active site, while the region in between is important for RBS binding and positions the TRS towards the active site (Hickman et al. 2004).

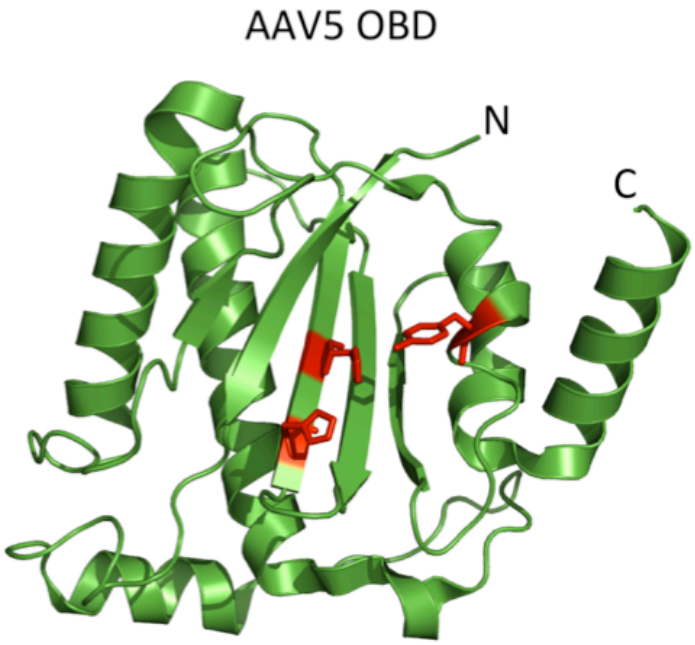


Figure 6: Crystal structure of the AAV5 Rep OBD.
The structure shows the five central anti-parallel strands, with the characteristic HUH motif located in the central strand and highlighted in red. This motif is facing the active site tyrosine located in one of the α-helices, also highlighted in red.

Like all HUH endonucleases, Rep catalyses ssDNA nicking and joining using the catalytically active tyrosine for a transesterification reaction that creates a 5'-phosphotyrosine intermediate and a free 3'-OH (Im and Muzyczka 1990; Snyder et al. 1990; Davis et al. 2000; Smith and Kotin 2000). In the case of AAV, the newly generated 3'-OH is used to prime replication of the ITR. Because the TRS-containing DNA in AAV is double-stranded, the endonuclease reaction must occur in two steps. The first step is the unwinding of the ITR to expose the TRS: this reaction is ATP dependent, requires the DNA helicase activity of Rep, and leads to the extrusion of a stem-loop structure exposing the nicking site in the loop (Brister and Muzyczka 1999; Davis et al. 2000). The second step is the actual nicking reaction, which is strand- and sequence-specific and requires a divalent metal ion, but is ATP independent (Brister and Muzyczka 1999; Smith and Kotin 2000). The TRS sequence recognised and cleaved by Rep is 3'-GGT/TGA-5' and is well conserved within AAV serotypes and *AAVS1* (Brister and Muzyczka 1999). As it is the case for the RBS sequence, a certain degree of flexibility is tolerated (Brister and Muzyczka 2000; Lamartina et al. 2000a). The only exception is the ITR of AAV5, in which both the TRS sequence and the distance from the RBS are different (Chiorini et al. 1999a). Unsurprisingly, AAV2 Rep supports the replication of other AAV serotypes except for that of AAV5 (Costello et al. 1997; Chiorini et al. 1999a).

1.2.1.4 Helicase domain

The helicase domain of Rep, shared by all the Rep isoforms, contains helicase and ATPase activities, as well as a nuclear localisation signal (Im and Muzyczka 1990; Yang et al. 1992; Smith and Kotin 1998; Cassell and Weitzman 2004). As explained previously, the helicase function of the large Rep proteins is necessary to allow efficient TRS nicking. In addition, Rep is capable of unwinding the ITR to allow re-initiation and the replicated dsDNA genome for continued replication *in vitro*, although it is possible that other helicases are involved in this process *in vivo*. Furthermore, the helicase activity of the small Rep proteins Rep52/40 is essential for the efficient packaging of AAV DNA.

Like all helicases, AAV Rep couples nucleotide hydrolysis to unidirectional translocation along nucleic acid molecules (Singleton et al. 2007; Patel et al. 2011). More specifically, Rep belongs to the superfamily 3 (SF3) of multimeric helicases which includes multifunctional proteins found mainly in small DNA and RNA viruses (reviewed in further detail below) (Gorbalenya and Koonin 1993; Hickman and Dyda 2005). This family of helicases has 3'→5' direction of unwinding and is characterised by four conserved motifs clustered in a short stretch of ~ 100 amino acids: motifs A and B (also known as Walker A and B boxes), motif B' (or the B' box), and motif C (or the sensor 1 motif). These motifs form the core active site, and are essential for nucleotide binding, metal coordination and for sensing of the phosphorylation state of the bound nucleotide (Hanson and Whiteheart 2005). The B' motif, unique to SF3 helicases, has been suggested to play a key role in DNA binding and in the coupling of ssDNA binding to ATP hydrolysis (Yoon-Robarts et al. 2004).

Structural studies of Rep40 (equivalent to the minimal helicase domain) revealed that it has a bi-modular architecture (Figure 7) (James et al. 2003; James et al. 2004). The N-terminus forms a small helical bundle of unknown function, while the C-terminal part is a larger α/β domain that contains the characteristic SF3 motifs (Figure 7). This α/β fold consists of a central 5-stranded antiparallel β sheet flanked on both sides by several α helices, and has the topology of AAA+ proteins (ATPases associated with diverse cellular activities) (James et al. 2003). AAA+ proteins are a vast family of P-loop ATPases that couple ATP hydrolysis to conformational changes and support a wide array of biological functions, including helicase activities such as that of AAV Rep (Iyer et al. 2004; Hanson and Whiteheart 2005). Rep, like other SF3 helicases, contains a modified version of the canonical AAA+ structure (James et al. 2003). In addition, it has an incomplete nucleotide binding site that requires the formation of an oligomeric interface to stabilise the binding of nucleotides through an arginine finger provided by a neighbouring subunit (James et al. 2004).

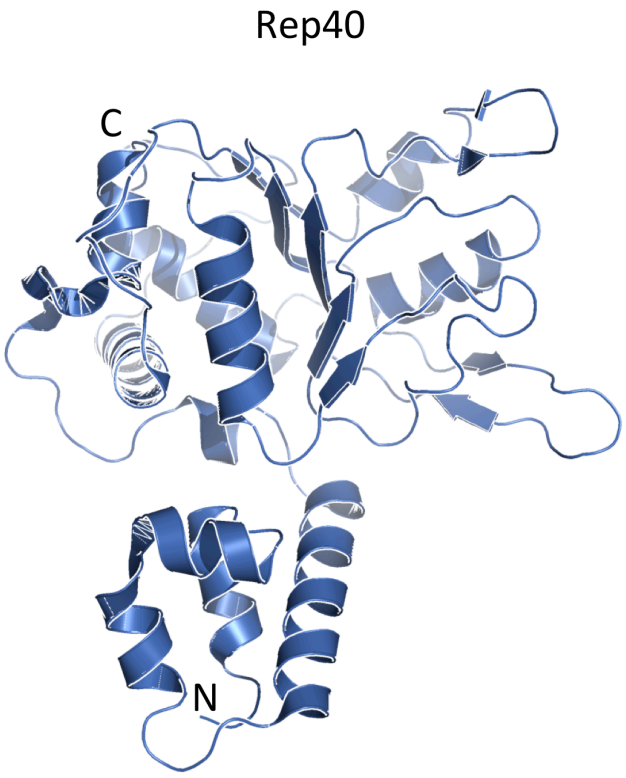


Figure 7: Crystal structure of Rep40.
The structure of Rep40, defined as the minimal helicase domain of AAV Rep, is bi-modular and contains an N-terminal small helical bundle (bottom) and an AAA+-like core characteristic of SF3 helicases.

1.2.1.5 Zn finger domain

Of the three domains of Rep, the C-terminal Zn finger domain, present only in the unspliced variants Rep78 and Rep52, is the least well characterised. While it has been shown that the Zn finger can bind zinc ions (Horer et al. 1995), its function and importance during the AAV life cycle are not clear. To date, the only functional differences identified between Rep78 and Rep68 and between Rep52 and Rep40 concern the interaction with other proteins (Di Pasquale and Stacey 1998) and the transcriptional regulation of heterologous promoters (Horer et al. 1995). Furthermore, the Zn finger domain has been suggested to be necessary for the AAV-induced de-regulation of the cell cycle by AAV (Berthet et al. 2005).

1.2.2 HUH endonucleases

HUH endonucleases are specialised enzymes that process site-specific ssDNA and catalyse cleavage and re-ligation reactions. They are widespread across the

three domains of life, and include Rep proteins involved in bacterial and viral rolling circle replication (RCR), relaxases that catalyse bacterial conjugation and transposases (Ilyina and Koonin 1992). This family of endonucleases is defined by the presence of two conserved protein motifs: the HUH motif, consisting in two histidines separated by a large hydrophobic residue, and the Y motif, containing one or two tyrosine residues and a lysine residue separated by several amino acids (Ilyina and Koonin 1992). HUH endonucleases can be classified based on the number of tyrosine residues: Y1 HUH endonucleases contain a single tyrosine residue in the Y motif, while Y2 proteins, which include AAV Rep78/68, contain two (Chandler et al. 2013). Surprisingly, although Rep78/68 contain two tyrosine residues – Y152 and Y156 – only Y156 is essential for catalysis (Davis et al. 2000). The ssDNA nicking mechanism involves the formation of a 5'-phosphotyrosine intermediate using one tyrosine from the Y motif and frees a 3'-OH at the cleavage site that can be used for various subsequent reactions, including the priming of replication. The HUH motif is necessary for the coordination of a divalent metal ion that facilitates the ssDNA cleavage reaction (Chandler et al. 2013). The relative position of the two motifs distinguishes the relaxases (or Mob) subgroup, in which the Y motif is found upstream of the HUH motif, and the Rep group that has the motifs arranged in the opposite order. Three-dimensionally, however, the arrangement of the two motifs relative to each other is identical (Figure 8). The HUH motif sits in the central β-strand of a four- or five-stranded antiparallel β-sheet, while the active site tyrosine(s) is positioned in a α-helix that faces the HUH motif (Figure 8) (Chandler et al. 2013). In addition to ssDNA nicking, many HUH endonucleases also recognise and bind DNA hairpin structures. Furthermore, in view of the functional variety of these proteins, it is not surprising that HUH domains are often combined with additional functional domains, notably helicase, oligomerisation and Zn-binding domains (Chandler et al. 2013). Finally, a few members of the HUH endonucleases family, including AAV Rep, the conjugative relaxase TrwC of plasmid R388 (Figure 8) and the TnpA(REP) transposase from *E.* coli K12, have been shown to possess the biotechnologically interesting ability to mediate site-specific recombination and integration into dsDNA (Gonzalez-Prieto et al. 2013).

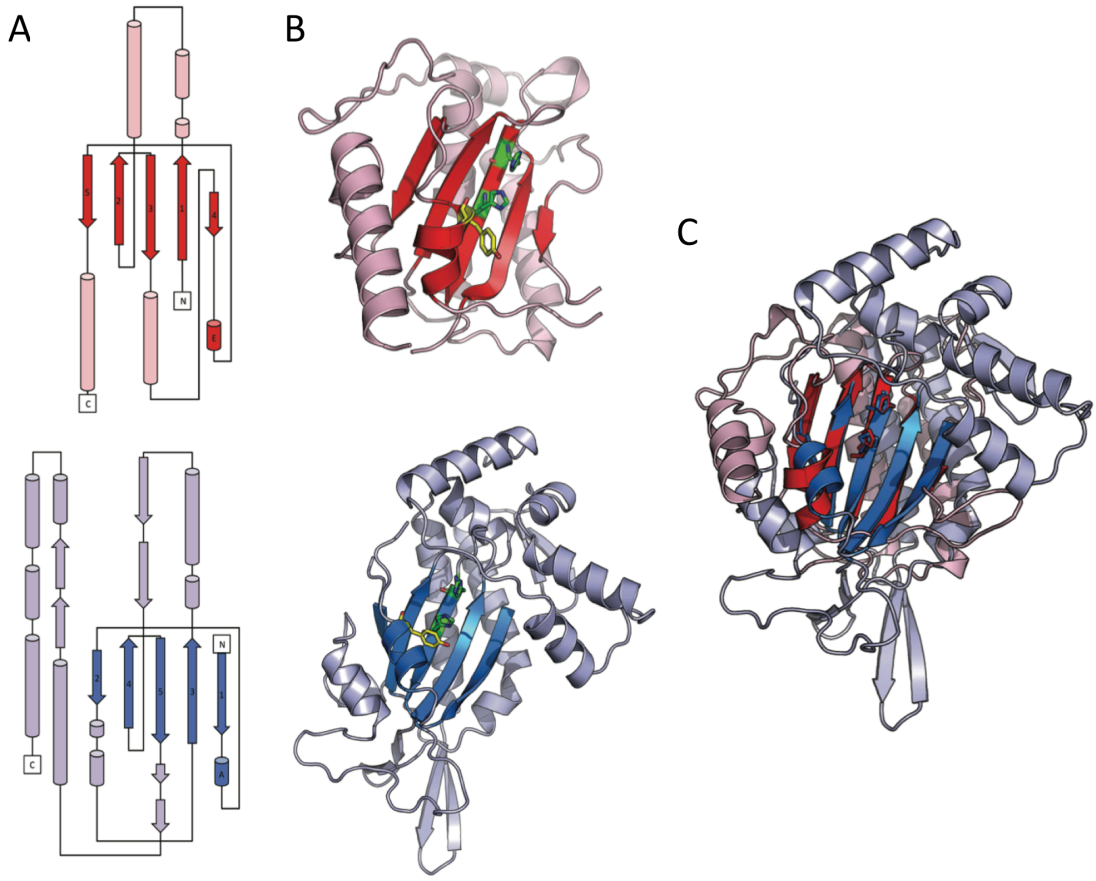


Figure 8: Structural conservation in HUH endonucleases.
(A) Topology diagram of the OBD of Rep (red, top) and TrwC (blue, bottom). The structures that form the core active site typical of HUH endonucleases are in dark red and dark blue, respectively. **(B)** Ribbon diagrams shown in the same orientation as in **(A),** with the active site tyrosine highlighted in yellow and the consensus HUH motif in green. **(C)** Structural alignment of Rep (red) and TrwC (blue) show a high degree of similarity between the two proteins active sites.

1.2.3 SF3 helicases

Helicases are molecular motors that use the energy obtained from NTP hydrolysis to translocate along DNA or RNA and unwind double-stranded regions. They are involved in every aspect of nucleic acid metabolism, including DNA replication, repair, transcription, translation and RNA metabolism. They form a vast class of enzymes that are identified and classified based on a series of conserved primary and tertiary structure motifs (Gorbalenya and Koonin 1993; Singleton et al. 2007). Helicase superfamilies 1 and 2 (SF1 and SF2) are the largest and best characterised; they contain seven conserved sequence motifs and include monomeric helicases that perform a wide variety of functions (Singleton et al.

2007). The oligomeric SF3 helicases have 3'-5' translocation directionality and group non-structural proteins found in the genomes of small DNA and RNA viruses (Gorbalenya and Koonin 1993). They are identified by only four conserved motifs (A, B, B' and C) and are topologically a modified version of AAA+ motor proteins (Hanson and Whiteheart 2005; Hickman and Dyda 2005). Motifs A and B are the canonical Walker A and B motifs, involved in ATP binding and divalent metal ion coordination, respectively. Motif C is the sensor 1 motif, which has been suggested to "sense" the phosphorylation state of the bound nucleotide. Motif B' binds DNA and is thought to be involved in the translation of ATP hydrolysis into the conformational changes that lead to translocation along the DNA (Yoon-Robarts et al. 2004). In addition, a conserved arginine finger in one subunit of the oligomeric complexes formed by these helicases contributes to ATP binding and hydrolysis in neighbouring subunits.

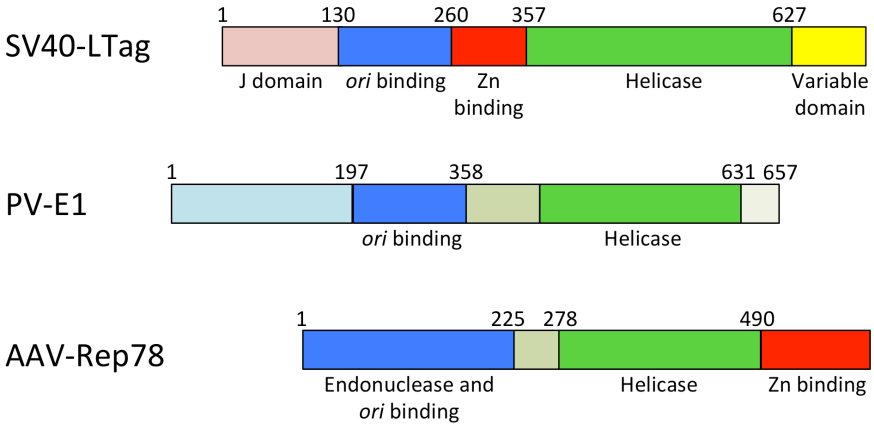


Figure 9: Domain architecture of SF3 helicases.
Schematic diagram representing the different domain organisations of three well-characterised SF3 helicases. The key functional domains, described throughout the text, are the helicase domain (in green) and the *ori* binding domain (in blue). Figure based on (Hickman and Dyda 2005).

The peculiarity of SF3 helicases is the combination of multiple enzymatic functions within a single protein. In particular, the highly conserved helicase domain is preceded by a less well-conserved origin-binding domain (Figure 9) (Hickman and Dyda 2005). This is especially relevant for small viruses that rely on the complex replicative machinery of the cell but need to recruit it to the viral

origin of replication. The best-characterised members of the SF3 helicases are the E1 protein of papilloma virus (PV), the large T antigen (LTag) of simian virus 40 (SV40) and AAV Rep (Figure 9); the crystal structures of both the origin binding and the helicase domains are available for all three proteins (James et al. 2003; Li et al. 2003a; Abbate et al. 2004; James et al. 2004). The core AAA+-like α/β fold is highly conserved amongst these three SF3 helicases, and consists of a 5-stranded anti-parallel β sheet flanked by α helices on both sides. Surprisingly, the overall fold of the origin binding domains is also conserved despite little sequence homology (Hickman and Dyda 2005). This is even more striking considering that the Rep OBD has an additional endonuclease function and binds direct repeats, in contrast to the palindromic sequences that are recognised by SV40-LTag and PV-E1. Furthermore, the genome structure and the replication strategies of SV40 and PV are very different from those of AAV. SV40 and PV genomes are covalently closed dsDNA circles and contain specific binding sites within the *ori* that are recognised by LTag and an E1/E2 complex, respectively. Binding is followed by the assembly of double-hexameric rings that leads to the unwinding of the *ori* and initiates bi-directional replication (Fouts et al. 1999; Gomez-Lorenzo et al. 2003; Li et al. 2003a; Abbate et al. 2004; Enemark and Joshua-Tor 2006; Fanning and Zhao 2009; Schuck and Stenlund 2011; Bergvall et al. 2013). In contrast, the AAV genome is linear and replication proceeds unidirectionally (Ward 2006), and the oligomerisation properties of Rep are still under debate (see section 1.2.4). Finally, sandwiched between the origin-binding and core helicase domains, is a small all-helical bundle of unknown function that is much less conserved. Structural studies of the hexameric rings formed by LTag and E1 revealed that this small domain is part of the oligomerisation interface (Li et al. 2003a; Enemark and Joshua-Tor 2006).

1.2.4 AAV Rep oligomerisation

Based on the fundamental role of oligomerisation in the mechanisms of origin melting and DNA unwinding described for the SV40-LTag and PV-E1 proteins, it is likely that oligomerisation also plays a critical role in the functions of the AAV Rep proteins. However, limited information is available on the oligomeric properties of

the AAV Rep proteins and the functional roles of the oligomeric complexes remain poorly characterised.

The small Rep proteins, Rep52 and Rep40, have been shown to be monomeric in solution (Smith and Kotin 1998; James et al. 2003; James et al. 2004; Dignam et al. 2007a; Mansilla-Soto et al. 2009). The fact that Rep52 and Rep40 are active helicases (Smith and Kotin 1998; Collaco et al. 2003), however, implies that they must exist at least as transient dimers to form an active ATPase site (James et al. 2004). The formation of dimers was indeed observed in the presence of ATP and ssDNA (Dignam et al. 2007a). In contrast to the monomeric behaviour of Rep40 the equivalent helicase domain of the SV40-LTag and PV-E1 proteins readily form hexamers (Sedman and Stenlund 1998; Li et al. 2003a). The structural determinants of these differences, however, remain unknown.

The oligomerisation of the large Rep proteins, which has mainly been studied using Rep68, is more complex, and contrasting results have been reported (Smith et al. 1997; Hickman et al. 2004; Dignam et al. 2007b; Mansilla-Soto et al. 2009). Rep68 has been suggested to form hexamers in the presence of the AAV *ori* in two separate reports (Smith et al. 1997; Dignam et al. 2007b), while two other studies have described five Rep68 molecules bound to an RBS-containing DNA and suggested a possible intermediate stage of assembly of a larger Rep complex on the AAV *ori* (Hickman et al. 2004; Mansilla-Soto et al. 2009). Mansilla-Soto and colleagues also described a double-octameric complex that is formed in the presence of a ssDNA substrate and is able to unwind a heteroduplex DNA substrate (Mansilla-Soto et al. 2009). In addition, the authors of this study provided evidence of substrate-dependent oligomerisation, by observing the formation of different oligomeric complexes in the presence of different DNA substrates (Mansilla-Soto et al. 2009). Taken together, these results suggest a complex and dynamic oligomerisation that may be directed by the DNA substrate interacting with Rep68. The functional importance of Rep78/68 oligomerisation and the properties that it confers to these multifunctional proteins remain to be addressed, and will be discussed in more details in Chapters 4 and 5.

CHAPTER 2: AIMS OF THIS STUDY

The aims of this study are **A.** to contribute to the understanding of the mechanisms of action of AAV Rep proteins by identifying functionally relevant structural determinants of AAV Rep68 oligomerisation; and **B.** to investigate the feasibility and the safety of Rep-mediated integration into the safe harbour locus *AAVS1* in a translationally relevant clinical model.

Accordingly, the main hypotheses established at the beginning of this thesis were:

- The substrate-directed oligomerisation of the remarkably multifunctional large Rep proteins of AAV contributes to the regulation of its enzymatic activities and coordinates its functions
- Rep-mediated integration into *AAVS1* is a naturally evolved strategy to site-specifically insert exogenous genetic material into the human genome that can be exploited for targeted transgene addition

All of the work presented in this thesis revolves around the multiple functions of the large Rep proteins of AAV. Result chapters 4 and 5 focus on the oligomeric properties of Rep68, and describe the identification of structural features that are essential for its oligomerisation and function. The data presented in the first two chapters are the product of a fruitful collaboration with the laboratory of Dr Carlos Escalante at the Molecular biophysics research centre of the Virginia Commonwealth University, which led to three publications (see annex 2) and an additional manuscript currently in preparation. The final results chapter presents a project that was designed to investigate the feasibility of using Rep-mediated targeted gene addition to modify cells with proliferative potential in a clinically relevant disease model, namely X-linked Severe Combined Immunodeficiency (SCID-X1). This chapter describes the generation and characterisation of SCID-X1 patient-derived induced pluripotent stem cells (iPSCs) as well as the generation and validation of the therapeutic targeting vector. Unfortunately, this study could not be completed due to an untreatable mycoplasma infection of the iPSCs.

CHAPTER 3: MATERIALS AND METHODS

3.1 MOLECULAR CLONING

3.1.1 Standard Polymerase Chain Reaction (PCR)

This is the basic PCR reaction that was used for colony screening and other tests. The polymerase kit used was the standard goTaq® DNA polymerase (Promega). PCR reactions were carried out in a total volume of 25µl containing (final concentrations): 200µM dNTPs (mix of dATP, dCTP, dGTP and dTTP, from NEB), 1X Colorless GoTaq Reaction Buffer, 1.25U GoTaq DNA polymerase, 1µM each of the primers, ~50ng DNA, and ddH_2O to a final 25µl volume. 0.01µmol of unmodified oligonucleotide primers for PCR amplification were obtained from Eurofins MWG Operon and re-suspended with ddH_2O to make stock solutions at 100µM. Forward (sense) and reverse (anti-sense, reverse complement) oligonucleotide primers were designed following the manufacturer's directions: 15-30 nucleotides in length, 40-60% GC content, and terminating in at least one G/C base. PCRs were performed using an Eppendorf Mastercycler ep gradient thermocycler. A typical PCR reaction initiated with a 2 minute denaturing step at 95˚C followed by 25-35 cycles including a denaturing step of 20 seconds at 95˚C, an annealing step of 45 seconds at 50°C (depending on the primers' melting temperature) and an extension step of 1minute/kb at 72°C. A final 5 minutes extension at 72°C followed the cycles. Reactions were kept at 4°C until use. For information on all the PCR primers used in this thesis please see Table 2 in Annex 1, p. 152.

3.1.2 High fidelity PCR

For all PCR reactions performed to amplify DNA fragments necessary for cloning, sequencing, and to prepare radiolabelled probes, the Phusion High Fidelity polymerase kit (Thermo Scientific) was used to ensure high fidelity amplification of target sequences. PCR reactions were carried out in a total volume of 25µl, in a mixture containing (final concentrations) 1-5ng of DNA template (more template was used for complex genomic DNA), 0.5µM of forward and reverse primers, 10mM

of dNTPs (NEB), 0.02u/µl Phusion® High-Fidelity DNA polymerase (Thermo Scientific), 1X Phusion® HF or GC Buffer and brought to the final volume with ddH_2O. PCRs were performed using an Eppendorf Mastercycler ep gradient thermocycler. Phusion GC buffer was used in reactions that initially did not work with HF buffer, as it is indicated for GC-rich templates or those prone to forming secondary structures. Reaction conditions were varied depending on the size and GC content of the fragment to be amplified. A typical PCR reaction initiated with a 30 seconds denaturing step at 98°C followed by 25-35 cycles including a denaturing step of 10 seconds at 98°C, an annealing step of 20 seconds at 60°C and an extension step of 30 seconds/kb at 72°C. A final 5 minutes extension at 72°C followed the cycles. Reactions were kept at 4°C until use. The annealing step temperature was adjusted to the melting temperature of the primers. Primers were designed and obtained from Eurofins MWG Operon as explained above. For some cloning strategies, primers were designed with external overhangs containing restriction sites matching those in the destination construct. For information on all the high fidelity PCR primers used in this thesis please see Table 2 in Annex 1, p. 153.

3.1.3 Site-directed mutagenesis

To generate Rep point mutants the Stratagene site-directed mutagenesis kit was used. Primers were designed to contain the desired point mutations in the centre flanked by 10-15 base pairs on either side matching the template sequence to allow sufficient annealing to the template DNA. This technology was used to introduce the following mutations in Rep: C151S, C405S, L193A-V196A, Y224A, Y224F, I251A. For information on all the mutagenesis primers used in this thesis please see Table 2 in Annex 1, p. 152.

3.1.4 Agarose gel electrophoresis

Amplified DNA PCR products or DNA fragments from restriction enzyme (RE) digestions from PCR were diluted with 5X DNA loading dye (NEB). Agarose gels were prepared by dissolving agarose powder (UltraPure™ Agarose, Life Technologies) in 1X TAE buffer (10X: Tris-Base 48.4g/L, Acetic acid 11.4ml/L, EDTA

3.7g/L) and heating until boiling. The agarose percentage varied between 0.7% and 1.5% depending on the size of the fragment to be observed/isolated, with higher percentages used to achieve better separation of smaller fragments. After allowing the dissolved agarose to cool down, 1µg/ml of Ethidium bromide solution (Sigma) was added and the mixture was poured into an electrophoresis tank (BioRad). Once solidified, the gel was immersed in 1X TAE buffer and DNA samples were loaded in the wells of the gel. Gels were run at 80-120V for about 30-90 minutes, depending on the separation required between bands. Band sizes were monitored with the 100bp or the 1kb DNA ladder (NEB), which was run in parallel. Ethidium bromide-stained DNA fragments were visualized under a ultra-violet trans-illuminator using a ChemiDocTM XRS+ System (BioRad).

3.1.5 Extraction and purification of DNA fragments from agarose gels

DNA fragments of interest were excised from agarose gels and purified using a QIAGEN Gel Extraction Kit. Three volumes of solubilisation and binding QG buffer were added to the agarose gel slice, the mixture was heated to 50°C for 10 minutes and vortexed to dissolve the slice. One volume of isopropanol was added to the mixture and subsequently added to a QIAquick spin-column. DNA was bound to the column by centrifugation at 10000rpm for 1 minute, after which the flow-through was discarded and the column washed with 750µl of ethanol-containing PE buffer and centrifuged twice to completely remove the ethanol. The DNA was finally eluted into a sterile eppendorf tube with 30µl of ddH_2O.

A similar procedure was used to purify RE digestion products that were not run on agarose gels. The same kit and protocol were used, with the exception that the RE digestion mixture was directly mixed with the QG buffer and isopropanol before proceeding with the column purification.

3.1.6 DNA digestion by restriction enzymes

For analytical RE digestions, 0.5µl of RE (NEB) was used to digest 500ng of plasmid DNA in a 1X solution of the buffer supplied by the manufacturer, and total reaction volume was adjusted to 20µl with ddH2O. For RE digestions to generate fragments for cloning, 1µl of RE was used to digest 2-5µg of DNA. All reactions were performed at the indicated temperature for the used RE (generally at 37˚C) for 2

hours. When digestions with two restriction enzymes were performed, a compatible reaction buffer was used as indicated by manufacturer. After single-enzyme digestions to generate plasmid fragments for cloning, vector DNA was treated with calf intestinal phosphatase (CIP, NEB) to remove the 5' phosphate group from the digested vector to prevent re-circularisation between compatible ends. 0.5µl of CIP were added directly to the RE digestion mixture and incubated at 37˚C for 30 minutes and further purified as described above.

3.1.7 DNA ligations

The purified DNA insert and vector were mixed at a 3:1 ratio with 1µl of 10X T4 ligase buffer (NEB) and 0.5µl of T4 ligase enzyme (NEB), in a total reaction volume adjusted to 10µl with ddH_2O. Reactions were incubated at room temperature for 1 hour or overnight at 16˚C.

3.1.8 Competent bacterial cells: media and maintenance

For DNA transformations and plasmid DNA amplification, two types of chemically competent cells were used. The *Escherichia coli* SURE supercompetent cells (Stratagene), that have reduced recombination potential, were used for transformation and amplification of plasmids containing AAV ITRs and grown at 30˚C. The *Escherichia coli* TOP10 competent cells (Life Technologies) were used for all other plasmids and cultured at 37˚C. Competent cells were stored at -80˚. Autoclaved LB (Fisher Scientific, 20g/L ddH_20) was used for liquid cultures. Autoclaved LB agar (37g/L ddH_20) set in Sterilin 10cm Petri dishes was used for solid phase growth cultures. Antibiotics were added to autoclaved broth or autoclaved agar cooled to 50°C. Ampicillin (Sigma, 100µg/ml in dH_2O) or Kanamycin sulfate (Fisher Scientific, 50µg/ml in ddH_2O) were used for selection. Stock preparations of transformed bacteria were kept as glycerol stocks (bacterial pellet resuspended in LB + 10-15% glycerol) at -80˚C.

3.1.9 Transformation of competent bacteria

5µl of the DNA insert-vector ligation reaction or 10-50ng of plasmid DNA was used for transformation into competent cells following the manufacturer's protocol. For SURE competent cells, 50µl of competent bacteria were initially

incubated with 1µl of β-mercaptoethanol for 10 minutes on ice. DNA was then added and incubated for 30 minutes on ice, followed by a heat shock at 42°C for 45 seconds and further 2 minutes on ice. For Top10 competent cells, 50µl of cells were mixed with DNA and incubated for 30 minutes on ice, followed by a 30 second heat shock at 42˚C and a final 2 minutes incubation on ice. In both cases, reactions were then incubated with 250µl of LB at 30˚C for 1 hour and plated on LB agar plates containing the appropriate antibiotic. Plates were incubated upside down at 30˚C or 37°C overnight.

3.1.10 Plasmid DNA amplification and purification – mini preps

Single colonies from transformed or re-streaked bacteria were picked and inoculated in 3ml of LB with antibiotic at the appropriate temperature overnight in an incubator shaker. The following day, 2ml of the culture were transferred to an eppendorf tube and cells were pelleted in a bench-top centrifuge at 8000rpm for 4 minutes. DNA was extracted using the QIAGEN mini prep kit. Briefly, cells were re-suspended in QIAGEN P1 buffer and then lysed in 100µl of P2 lysis buffer and mixed by inverting the tube. After a 5-minutes incubation at room temperature, chilled P3 neutralisation buffer was added to neutralise the mixture and samples were incubated on ice for 5 minutes. DNA was separated from bacterial debris by centrifugation at 13000rpm for 10 minutes in a bench-top centrifuge. Supernatant containing DNA was then transferred to a QIAprep spin column to bind DNA followed by washing with ethanol containing PE buffer and elution in 40-50µl of ddH_2O. DNA was kept at 4˚C for short term storage or at -20˚C for longer storage.

3.1.11 Plasmid DNA amplification and purification – midi/maxi preps

1ml of transformed bacteria culture or 10µl from a glycerol stock were inoculated overnight in 100ml (300-500ml for maxi preps) of LB with antibiotic selection. Cells were harvested by centrifugation at 6000rpm for 10 minutes at 4°C and DNA extraction was performed using the NucleoBond® Xtra Midi/Maxi kit (Macherey-Nagel), following the manufacturer's protocol. The general principle behind this plasmid purification protocol is based on a modified alkaline lysis procedure, followed by plasmid DNA binding to a NucleoBond resin under appropriate low salt and pH conditions. RNA, proteins, dyes and low molecular–

weight impurities are removed by several wash steps, and the plasmid DNA is finally eluted in a high-salt buffer, concentrated and desalted by isopropanol precipitation, and washed in ethanol. Finally, the clean DNA pellet was allowed to air dry before re-suspending in 100-1000µl ddH_2O depending on the amount of purified DNA and the desired final concentration. DNA concentration was subsequently measured (see below). Preparations obtained by this method were typically at a concentration of 1µg/µl and stored short-term at 4˚C or long-term at -20˚C.

3.1.12 Determination of DNA concentration and DNA sequencing

DNA concentration was determined using a Nanodrop ND-100 Spectrophotometer (Labtech International). Following a blank measurement, 1µl of undiluted DNA preparation was loaded onto the measuring pedestal. The Nanodrop calculates the DNA concentration measuring the sample absorbance at 260nm (OD_{260}) and assuming that 1 OD_{260} unit corresponds to 50µg/ml of dsDNA. The purity of the DNA sample can be assessed by the OD_{260}/OD_{280} ratio, which at ≈1.8 is considered 'pure' for DNA (free from protein contamination).

DNA sequencing was performed at Eurofins MWG Operon from 1.5µg of plasmid DNA in 15µl ddH_2O pre-mixed with 2µl of the appropriate primer at 10pmol/µl. Sequencing was performed using the Value Read service in tube format, results were returned on-line and subsequently analysed using the APE DNA analysis software and the NCBI Blast tool. Information on the primers used for sequencing can be found in Table 2, Annex 1, p. 152.

3.1.13 Plasmids

Here, I present the basic plasmids that I have used during my thesis and those that have required cloning. When the cloning strategy is identical for several plasmids, the general strategy is described. For a list of all the plasmids used, please see Table 1 in annex 1, p. 150.

3.1.13.1 WT AAV and Rep68 plasmids

The pDG plasmid is a valuable tool for the production of recombinant and wt AAV. It contains all the adenovirus helper factors necessary for AAV production, as well as the AAV2 Rep and Cap ORFs (Grimm et al. 1998).

The mini-pDG plasmid was derived by Dr Els Henckaerts from pDG and contains only the AAV2 Rep and Cap ORFs. All the mutagenesis to generate the Rep mutants were performed on this plasmid.

The pAV2 plasmid is infectious and contains the full WT AAV genome inserted between two BglII linkers (Laughlin et al. 1983). All the Rep mutants were cloned into this plasmid using a SfiI and HindIII fragment generated from the mini-pDG variants.

pCMV(Δ6)-Rep68 was derived from the pCMV(Δ6) plasmid, kindly provided by Prof Martin-Serrano, part of a series of plasmids containing truncated versions of the CMV promoter designed to fine-tune the level of expression of a protein of interest (Morita et al. 2012). The pCMV(Δ6) plasmid was selected for the low levels of Rep68 obtained upon 293T cells transfection. The Rep68 sequence was derived from the pRep68-M225G plasmid (cloned by Dr Nathalie Dutheil in the lab) by PCR and cloned into the pCMV(Δ6) multiple-cloning site using PCR overhangs and the REs XbaI and XhoI. All the Rep mutants were cloned into the WT pCMV(Δ6)-Rep68 using a SfiI and HindIII fragment generated from the mini-pDG variants.

3.1.13.2 rAAV production plasmids

The pTRUF11 plasmid carries the humanised green fluorescent protein (hGFP) sequence under the control of the hybrid CMVie enhancer/chicken β-actin promoter (CAG) and the SV40 early polyA signal, together with the neomycin resistance gene driven by the HSV thymidine kinase promoter and the bGH polyA, flanked by the AAV2 ITRs. This plasmid is used to produce rAAV2-GFP (Zolotukhin et al. 1996).

The pHGTI-Adeno1 plasmid contains all the adenovirus helper functions necessary for AAV production and is used to produce WT or recombinant AAV by double or triple transfection schemes, respectively (Streck et al. 2006).

The *IL2RG*-containing plasmids were generated by cloning an *IL2RG*-containing cassette from the SINLV-IL2RG vector (kindly provided by A. Thrasher, Institute of Child Health, London, UK). This vector contains a short version of the Elongation Factor 1 alpha promoter that drives the expression the *IL2RG* gene. This expression cassette was cloned into the pTRUF11 ITR-containing plasmid using XhoI and SacI to generate the pTRUF11-GFP-CG vector. To generate the pUTRF11-CG construct, the GFP expression cassette from pTRUF11-GFP-CG was removed using KpnI and ClaI. The other *IL2RG*-expressing plasmids were generated in the pSUB201 backbone, containing modified ITRs for reduced recombination and the WT AAV genome flanked by XbaI sites (Samulski et al. 1987). The AAV genes were replaced by the *IL2RG* expression cassette using the XbaI sites to generate pMB13-CG, and additional restriction sites were added on both sides of the cassette to allow for further modifications. pSUB201-IL2RG was generated by adding a WPRE element and a new polyA (SV40 early) obtained by PCR to pMB13-CG using XhoI and AvrII sites. To generate the pSUB201-IL2RG-GFP construct, a CMV-GFP-bGHpolyA expression cassette was cloned using XhoI and NotI sites into pSUB201-IL2RG.

3.2 CELL CULTURE AND TRANSFECTIONS

3.2.1 Cell lines, media and maintenance

All cell culture work was performed in assigned mycoplasma-free tissue culture hoods. Adherent HeLa and 293T cells were obtained from the American Tissue Culture Collection (ATCC). HeLa cells derive from a human epithelial cervical adenocarcinoma (initially from a patient, Henrietta Lacks), while HEK293T cells are human embryonic kidney cells modified to contain the SV40 large T antigen. These cells were grown in Dulbecco's Modified Eagle Medium (DMEM) +4.5g/L Glucose, +L- Glutamine, +Pyruvate (GIBCO 41966, Invitrogen), and 10% heat inactivated (56°C for 30 minutes) foetal bovine serum (FBS, Invitrogen). Cells were cultured at 37°C and 5% CO_2 in 10cm dishes (Corning). Unless otherwise required, cells were regularly passaged using 1ml TrypLE^{TM} Express (GIBCO 12605, Invitrogen). To passage them, cells were incubated with trypsin for 1-3 minutes at 37°C, re-

suspended by repeated pipetting and the required proportion was transferred to a 10cm dish containing pre-warmed fresh complete DMEM.

ED7R cells are a human T-cell line that does not express the common γ-chain (Zhang et al. 2007) and were kindly provided by Prof Adrian Thrasher. ED7R-CG cells are identical to ED7R cells but contain 3 integrated copies of the *IL2RG* gene. ED7R and ED7R-CG cells grow in suspension and were cultured in RPMI 1640 (GIBCO, Invitrogen) medium at 37°C and 5% CO_2 in T25 flasks (Corning). Unless otherwise required, cells were regularly passaged by dilution.

3.2.2 Freezing and thawing of cell lines

Stocks of frozen cells were long-term stored in liquid nitrogen. For freezing cells, at least 2-3 confluent 10cm dishes of each cell line were prepared. Cells were detached using 1ml of trypsin and washed in 5-10ml fresh culture medium. After harvesting the cells by centrifugation at 1200rpm for 5 minutes, cells were re-suspended in 1ml of freezing medium containing 10% DMSO (Sigma), 50% FBS (Invitrogen) and 40% culture medium and immediately transferred to a 1.8ml labelled cryovial. Cryovials were initially stored at -80˚C and after 24-48 hours the cryovials were moved to a liquid nitrogen tank. To later recover the cryo-preserved cells, vials were thawed rapidly at 37°C, cells were pelleted at 1200rpm for 5 minutes and rapidly re-suspended in 10ml of pre-warmed DMEM with 10% FBS. Cells were left to recover in 10cm dish for at least 24-48 hours.

3.2.3 Mycoplasma testing

Standard mycoplasma testing was performed regularly on all cells cultured in the mycoplasma-free tissue culture room using the MycoAlert™ Mycoplasma Detection Kit (Lonza) and the controls from the MycoAlert™ Assay Control Set (Lonza) following the manufacturer's protocol. This test is based on the activity of mycoplasmal enzymes that are found in all six of the main mycoplasma cell culture contaminants but are not present in eukaryotic cells, which convert ADP present in the MycoAlert substrate to ATP. The ATP is then transferred into a light signal via the luciferase enzyme in the MycoAlert Reagent. By measuring the level of ATP in a sample both before (read A) and after the addition of the MycoAlert Substrate

(read B), a ratio can be obtained which is indicative of the presence or absence of mycoplasma.

3.2.4 Transient transfection of DNA plasmids

For regular plasmid DNA transfection, 293T cells were seeded 24 hours beforehand and transfected using polyethylinimine (PEI, Polysciences). For a 24-well plate, 1µg of plasmid DNA was combined with 4µl of PEI in 50µl serum-free DMEM for 10 minutes before drop-wise addition to cells with fresh DMEM. For transfections that required high efficiency, cells were seeded 24 hours beforehand and transfected using Lipofectamine 2000 (Invitrogen) according to manufacturer's instructions. For a 24-well plate, 1µl Lipofectamine was mixed with 50µl Optimem (GIBCO 31985, Invitrogen) per sample, mixed and incubated at room temperature for 5 minutes. In a separate eppendorf tube, 50µl of Optimem were added to 1µg of plasmid DNA. The transfection reagent mix was then combined with the DNA mixture and the samples were incubated at room temperature for 20 minutes before drop-wise application to the cells. In all cases, the culture media was changed 4-8 hours post transfection. For larger-scale transfections, quantities were scaled up accordingly.

3.3 AAV PRODUCTION, PURIFICATION AND QUANTIFICATION

3.3.1 AAV production

Virus productions carried out during this thesis were relatively small. 2X T225 flasks were seeded with 293T cells. The following day the cells were co-transfected using PEI: 13.8µg of ITR-containing plasmid, 13.8µg of pDG and 41.5µg of HGTI (1:1:1 molecular ratio) were mixed into 5ml serum-free DMEM, and combined with 5ml serum-free DMEM containing 242µl PEI (1mg/ml, pH 7.0, use 3.5ml PEI per mg of DNA). The DNA/PEI/DMEM mixture was incubated 20 minutes and then added to 80ml pre-warmed complete DMEM and added to the cells. For WT AAV (or mutants) production, only the pAV2 plasmid was transfected (without pDG). After three days, the cells were detached by tapping/shaking the flasks, and transferred to a 50 ml tube per flask. The cells were spun at 1200rpm for 10 minutes, washed with PBS and spun again before being lysed in 150 mM NaCl, 50 mM Tris at pH 8,5,

followed by three freeze (-80˚C) - thaw (37˚C water bath) cycles. The lysate was then treated for 30 minutes at 37˚C with 150 units/ml of benzonase (Sigma, requires a working concentration of 1mM $MgCl_2$), to digest all DNA and/or RNA not encapsidated, and subsequently cleared by centrifugation. The lysate was then used to purify AAV by the iodixanol gradient method (see below).

3.3.2 AAV purification by iodixanol density gradient

The lysate was overlaid on top of an iodixanol (Optiprep® Density Gradient medium, Sigma) gradient, and then spun at 40000rpm for 3h at 18˚C (in a Sorvall discovery 90SE ultracentrifuge, TH641 rotor), to separate the components of the lysate. AAV particles are found at the 40-60% iodixanol interface after the ultracentrifugation from were it is extracted using a syringe. The following density fractions form the gradient (bottom-to-top): 1.55ml 60% (1.55ml Optiprep, 3.88µl phenol red), 1.55ml 40% (1.05ml Optiprep, 0.31ml 5X TD, 0.186ml ddH_2O), 1.88ml 25% (0.78ml Optiprep, 0.376ml 5X TD, 0.724ml ddH_2O, 4.7µl phenol red), 2.8ml 15% (0.7ml Optiprep, 0.56ml 5M NaCl, 0.56ml 5X TD, 0.98ml ddH_2O). The 5X TD buffer contains 5X PBS, 5mM $MgCl_2$ and 12.5mM KCl. To re-buffer the virus, samples (the iodixanol layer containing AAV particles) were diluted in Lactated Ringer's solution (Hartmann's solution, Baxter), added to a sterile Vivaspin 20 100KDa cut off concentrator (Sartorius Stedium Biotech), and spun down to a final volume between 500 and 800µl. The viral prep was finally aliquoted and aliquots were stored at -80˚C for long-term storage or at 4˚C for immediate use.

3.3.3 Quantification of viral particles

The quantification of AAV particles was performed as previously described (Kohlbrenner et al. 2012). A BSA standard curve and purified AAV samples were loaded and ran on a polyacrylamide gel (see section 3.4.1 for details on SDS-PAGE). The BSA standard curve was prepared using at least 5 BSA dilutions, ranging between 2000ng to 31.25ng, 10µl of appropriate BSA dilution was mixed with 5µl of water and 3µl of 6X loading buffer (0.8g SDS, 5ml Tris pH 6.8, 5ml glycerol, 4mg bromophenol blue, 5% beta-mercaptoethanol). For viral samples, 15µl of purified virus were mixed with 3µl of 6X loading buffer. The gel was migrated at 80V until the blue stain ran out of the gel. The gel was then fixed with a solution of 50%

ethanol and 15% acetic acid in water, washed with water, and stained with the Krypton™ Infrared Protein Stain. The destaining solution was made with 5% acetic acid and 0.1% Tween®-20 in water. The gel was finally imaged with an Odyssey® Infrared Imaging system (Li-Cor), and quantified with the Odyssey software. The protein content (in ng) of the band corresponding to the viral VP3 protein was deduced from the BSA standard curve, and the viral particle titre was calculated from the known amount of VP3 proteins per capsid ($4.986868*10^{-9}$ng/capsid) and the volume of virus loaded.

3.3.4 Quantification of viral genomes

The viral genome titre (or genome-containing particle titre) was quantified by quantitative PCR (qPCR) using the absolute quantification method described in section 3.5.3. Primers were designed (see Table 2, Annex 1, p. 152 for details) for all AAVs produced (WT or recombinant) to be used at the standard conditions described in section 3.5.3.

3.3.5 AAV infection and transduction

The same basic principle applies to all infections (using WT or mutant AAVs) and transductions (using rAAVs) performed during this thesis. Briefly, cells were infected/transduced at 60/70% confluency in low volume (200µl for a well of 24-well plate, volumes adjusted proportionally to the size of the culture vessel) and incubated for 1h at 37˚C while gently shaking the plate/dish every 15 minutes. This adsorption step ensures maximal infection/transduction efficiency. After 1h, the medium was replenished to the normal culture volume (i.e. 500µl for a well of 24-well plate). Medium was replaced after 24 hours. Cells were generally harvested after 48-72 hours for further analysis (Western blot, section 3.6.2, or flow cytometry, section 3.6.4).

3.4 DNA, RNA AND PROTEIN EXTRACTION FROM EUKARYOTIC CELLS

For some of the results presented in Chapter 4, DNA, mRNA and protein were extracted from cells derived from a single transiently transfected and adenovirus-infected 10cm dish split in 4.

3.4.1 Total DNA extraction

For total/genomic DNA extraction the QIAGEN DNeasy Blood & Tissue Kit was used as per manufacturer's instructions. Cells were first washed in PBS and pelleted, followed by lysis and proteinase K digestion to remove all proteins. The DNA isolation/purification was performed using a silica-based DNA purification in spin-columns using the kit buffers.

3.4.2 Isolation of small molecular weight DNA (Hirt extracts)

For specific isolation of small molecular weight DNA a modified version of the Hirt extraction protocol (Hirt 1967) was used. The volumes presented here are for extraction of DNA from 293T cells from a ¼ of a 10cm dish. Cells were harvested and pelleted in their medium, and subsequently washed in PBS. Cells were pelleted by centrifugation for 5 minutes at 1200rpm, then resuspended in 50µl PBS and transferred to a clean eppendorf tube. Cells were lysed in 500µl Hirt buffer (0.6% SDS, 10mM Tris HCl pH 7.4, 10mM EDTA), followed by lysate digestion using 50µg/ml proteinase K (Sigma) for 1 hour at 37˚C. 120µl 5M NaCl was added (final concentration 1M), the tube was gently mixed and stored overnight at 4˚C. The following day, the tubes were spun for 45 minutes at 4˚C at maximum speed on a bench-top centrifuge (13000rpm) and the supernatant (approx. 500µl) carefully collected in a new tube. This was followed by addition of 1 volume (500µl) of Phenol/Chloroform/Isoamylalcohol (25:24:1 mixture, Sigma), gentle mixing, a 10 minutes spin at 4˚C at maximum speed on a bench-top centrifuge (13000rpm), transfer of the supernatant (approx. 400µl) carefully collected in a new tube, addition of 40µl 3M Na acetate (pH 5.5, final concentration 0.3M) and 0.7 volumes (350µl) isopropanol and gentle mixing. The sample was stored at -20˚C overnight for DNA precipitation. On the final day, the mixture was spun 20 minutes at 4˚C at maximum speed in a bench-top centrifuge (13000rpm), the supernatant discarded and the pellet washed in 70% ethanol before air drying in a laminar-flow cabinet. The DNA pellet was finally resuspended in 50µl ddH_2O containing 100µg/ml RNAse (Sigma), and stored at -20˚C long-term.

3.4.3 Total RNA extraction

RNA extraction was performed using the RNeasy mini Kit (QIAGEN), following the manufacturer's instructions. Cells were first washed in PBS and pelleted, followed by lysis in the supplied buffer, homogenisation using a syringe and a 21G needle, and column purification. RNA was eluted using 50µl RNAse-free water (supplied) and the concentration determined using the Nanodrop spectrophotometer in a procedure analogous to that used for DNA (see section 3.1.12). At this point, 10µg of RNA were diluted in a final volume of 87.5µl RNAse-free water and mixed with 10µl Buffer RDD (Quiagen DNAseI kit) and 2.5µl of DNaseI stock solution (Quiagen DNAse I kit) and incubated for 20 minutes at room temperature. The DNAse I-treated RNA was then re-cleaned using the RNeasy kit (following the RNA clean-up protocol, supplied). The DNaseI stock solution was prepared by dissolving the lyophilized DNaseI (1500 Kunitz units, QIAGEN) in 550µl of the RNase-free water provided. RNA was stored at -80˚C.

3.4.4 Total cellular protein extraction

Volumes indicated are for a well of a 24 well plate. Cells were harvested and washed in PBS and pelleted. The cell pellet was then dissolved in 50µl RIPA buffer (25mM Tris HCl pH 7.4, 150mM NaCl, 1% NP-40, 1% sodium deoxycholate, 0.1% SDS) containing 2µl of 25X protease inhibitor cocktail (Complete, Roche), and incubated on ice for 15 minutes. The lysate was then spun at 13000rpm for 10 minutes at 4˚C, and the supernatant transferred to a clean 1.5ml tube. Proteins were stored at -80˚C.

3.5 DNA AND RNA DETECTION AND ANALYSIS

3.5.1 Preparation and detection of radiolabelled probes

For this study, two radiolabelled probes were used: one for AAV Rep and the other for the ampicillin gene present in our plasmids. For the Rep probe, primers ND44 and ND45 were used, while for the ampicillin probe primers MB5 and MB6 were used (see Table 2 in Annex 1, p. 152, for sequences). Primers were used to amplify the probe sequence from a plasmid, and the amplified PCR product was cloned into the TOPO vector using the TOPO-TA cloning system (Invitrogen). The

TOPO-probe vector was used from a second PCR to amplify the probe, which was then cleaned and labelled with [^{32}P]dCTPs (Perkin Elmer) using the Prime-It RmT random primer labeling kit (Stratagene), following the manufacturer's instructions, [^{32}P]dCTPs. UV cross-linked membranes from slot blot or Southern blot experiments (see sections 2.5.2 and 2.5.3, respectively) were prehybridised in 0.75X nylon wash solution (NW; 40.6g Na_2HPO_4, 18.65g EDTA, 500g SDS in 3.58 litres of ddH2O, pH 7.2) buffer at 65°C. The membranes were hybridised overnight in 0.75X NW buffer to either of the radiolabeled probe. The membranes were then washed twice with 0.5X NW buffer, followed by an additional wash in 0.1X NW buffer. Finally, the membranes were exposed to a phosphor screen (GE Healthcare Life Sciences) for 2 hours up to overnight (depending on the strength of the signal). Images were acquired using the Typhoon imaging system (GE Healthcare Life Sciences) and analysed with the ImageQuant software (GE Healthcare Life Sciences).

3.5.2 Slot blot

293T cells were transfected with Lipofectamine 2000 (Invitrogen, see section 3.2.4) with the pAV2 (or mutants) plasmid and after 24h infected with adenovirus at an MOI of 10. After an additional 48h, the cells were harvested, lysed in 350µl of 0.2M NaOH–10mM EDTA, and boiled for 15 minutes. Each sample was loaded in 100µl triplicates onto a nylon hybridisation membrane (Amersham Biosciences) using a Bio Dot apparatus (Biorad). The membranes were rinsed in 2X SSC (1X SSC is 150mM NaCl plus 15mM Na citrate, pH 7), air-dried, UV cross-linked, and radiolabelled as explained above.

3.5.3 Southern blot

The samples used for Southern blots presented in this thesis were obtained by Hirt extracts (section 3.4.2) and used to study AAV replication (section 3.7.2). A 300ml 0.8% agarose gel was prepared as described in section 3.1.4. Samples were migrated overnight at 35V. Then, the gel was incubated 30 minutes rocking in denaturing solution (1.5M NaCl, 0.5M NaOH in ddH_2O), washed shortly in ddH_2O twice, and incubated twice 30 minutes in neutralising solution (Tris 0.5M pH7.4, 1.5M NaCl in ddH_2O). The transfer system was then assembled: the gel was placed face down on a Whatman paper soaking in 20X SSC (175.3g NaCl, 88.2g Na citrate in

1l ddH_2O), topped with a nitrocellulose membrane (GE Healthcare Life Sciences), 2 layers of Whatman paper the same size of the gel, several tissues for adsorption and finally 2 weights. The transfer of the DNA from the gel to the nitrocellulose membrane occurs by capillarity, and is done overnight. The following day, the membrane was soaked in 2X SSC, air-dried, UV cross-linked and radiolabelled as explained in section 3.5.1.

3.5.4 qPCR: absolute quantification

Absolute quantification of AAV genomes was performed by qPCR using the SYBR Green Jump Start Taq ready mix without $MgCl_2$ (Sigma) according to the manufacturer's instructions on an ABI Prism cycler (AB Applied Biosystems). The $MgCl_2$ and primer concentrations were optimized previously (Zeltner et al. 2010), to 4mM and 0.25µM respectively. Linearised mini-pDG plasmid was used to prepare the standard curve for WT quantification, while pMB13 was used for the *IL2RG* standard curve. Standard curves had concentrations ranging from 2ng (corresponding to 2.53×10^8 molecules of dsDNA) to 0.2ng (corresponding to 2.53×10^5 molecules of dsDNA). 1µl of purified and concentrated virus and three 10-fold serial dilutions were used as viral samples. The primers used were Cap1 and Cap2 for WT AAV quantification, CGfw and CGrv for quantification of the *IL2RG* vectors. The cycling parameters were the following: 2 minutes at 94˚C followed by 40 times 15 seconds at 94˚C, 15 seconds at 58˚C and 1 minute at 72˚C. The amount of genomes in the viral samples was calculated from the amount of DNA detected. Each standard, control and sample was loaded in duplicate. For details on all primers see Table 2 in Annex 1, p. 152.

3.5.5 Reverse transcription

For RNA analysis, RNA was reverse-transcribed using the High-Capacity cDNA Reverse Transcription Kit (Applied Biosystems) following the manufacturer's instructions. First, a 2X RT master mix was prepared by mixing (volumes for one reaction) the following provided reagents: 2µl 10X RT buffer, 0.8µl dNTP mix (100mM), 2µl 10X RT random primers, 1µl MultiScribe™ Reverse Transcriptase, 4.2µl nuclease-free ddH_2O. 10µl of 2X RT master mix was then mixed with 10µl RNA at the desired concentration. Reverse transcription was performed using a thermal

cycler with the following programme: 10 minutes at 25˚C, 2 hours at 37˚C, 5 minutes at 85˚C. The obtained cDNA was stored at 4˚C for short-term storage or at -20˚C for long-term storage, and was used for relative quantification by qPCR (see below).

3.5.6 qPCR: relative quantification

3'FAM-5'TAMRA-conjugated qPCR probes and qPCR primers designed to bind into the Rep or the Cap genes were obtained from MWG Eurofins (see Table 2, Annex 1, p. 152, for details). qPCR was performed on 50ng cDNA using TaqMan Universal PCR master mix (Life Technologies) and the custom primer-probe mix of choice, primers were used at a final concentration of 900nM, and probes at 250mM. Relative expression levels were determined by the comparative threshold cycle (Ct) method (Schmittgen and Livak 2008).

3.6 PROTEIN DETECTION AND ANALYSIS

3.6.1 Sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE)

4-10µl of protein extracts in RIPA buffer (section2.4.4) were mixed with 1-2µl of loading buffer (0.8g SDS, 5ml Tris pH 6.8, 5ml glycerol, 4mg bromophenol blue, 5% beta-mercaptoethanol) and denatured by boiling for 5 minutes. Subsequently, proteins were separated in 10% or 12% polyacrylamide mini-gels. The percentage was determined by altering the volume of 40% Acrylamide solution (Acrylamide:Bis-Acrylamide 29:1, Fisher Scientific) and ddH_2O to a reaction mixture of 1.5M Tris-HCl pH8.8, 0.4% SDS, 3.3µl/ml 10% ammonium persulphate (APS, Sigma) and 0.6µl/ml of N,N,N'N'- Tetramethylethylenediamine (TEMED, Sigma). The stacking gels were made using 0.5M Tris-Cl pH6.8, 0.4% SDS, 5µl/ml 10% APS, 1µl/ml TEMED and ddH_2O. Polymerised gels were transferred to the running tank and immersed in running buffer (100ml 10X Tris-glycine and 10ml 10% SDS and ddH_2O to 1l). 10X Tris-glycine was prepared by mixing 61g Tris base and 288g glycine in 2l ddH_2O. Gels were typically run at 120V until the blue dye front had migrated to the bottom of the gel, using the Protean II mini gel electrophoresis kit (BioRad). 2µl of Precision

Plus Protein Dual Color Standards (Bio-Rad) was used as a guide for molecular weights in all gels.

3.6.2 Western blot

Proteins resolved in SDS-PAGE were transferred onto pure nitrocellulose 0.45µm membrane (VWR International). Gel to membrane transfer was achieved using a BioRad Mini Trans-Blot® transfer system with transfer buffer (100ml 10X Tris-glycine, 200ml methanol, 700ml ddH_2O) at 16V overnight or at 100V on ice for 75 minutes. Membranes were then removed from the transfer tank and blocked with 5% (w/v) skimmed dried milk in PBS-Tween (1X PBS, 0.1% Tween 20, Sigma) for 30 minutes. Membranes were then incubated with the indicated primary antibody in 5% milk in PBS-Tween for 3 hours at room temperature or overnight at 4˚C. Next, membranes were washed 3 times for 15 minutes in PBS-Tween and consequently incubated with the corresponding secondary antibody for 1 hour. Finally, membranes were washed three times for 15 minutes in PBS-Tween and developed according to the antibody and technology used (see section below).

3.6.3 Protein detection methods

Infrared IRDye®-conjugated 680nm and 800nm secondary antibodies were detected using a LI-COR Odyssey infrared scanner (LI-COR Biosciences). Membranes were placed on the scanner with a few drops of 1X PBS to avoid them drying out but making sure there were no air bubbles between the membrane and the scanning surface. Laser intensities were adjusted according to the signal obtained.

Alternatively, the signal of membranes incubated with horseradish peroxidase (HRP)-conjugated secondary antibodies was detected with a SuperSignal West Pico chemiluminescene system (Thermo Scientific), prepared according to the manufacturers' instructions and applied on the membrane for 5 minutes. The membrane was then sandwiched between two clear plastic sheets and visualised using a Image Quant LAS4000 (GE Healthcare) chemiluminescence imaging technology.

Information on all the primary and secondary antibodies that were used to generate the data presented in this thesis can be found in Table 3, Annex 1, p. 154.

3.6.4 Flow cytometry

GFP transfected or transduced cells were washed, passed through a cell strainer to ensure single-cell separation, and analysed for protein expression by flow cytometry on a BD FACSCantoII flow cytometer using the BD FACSAriaII program. All results obtained by flow cytometry were analysed using the FlowJo software. Information on all the FACS antibodies that were used to generate the data presented in this thesis can be found in Table 3, Annex 1, p. 154.

3.6.5 Protein structure analysis

All the protein structures presented were generated using the Mac PyMOL protein visualisation software.

3.7 IMAGING

3.7.1 Cell fixation and immunostaining

iPSCs were prepared in 24-well plates. Following aspiration of the medium, cells were washed once with 1X PBS and fixed (protein cross-linking) in 4% paraformaldehyde (PFA, Thermo Scientific, diluted in PBS, pH 7.4) for 15-20 minutes at room temperature. The cells were then washed twice in PBST (0.05% Tween-20 in 1X PBS), permeabilised in 0.1% Triton X-100 (Sigma) in 1X PBS for 10 minutes, washed twice again in PBST, blocked in blocking solution (4% FBS in 1X PBS) for 30 minutes, stained for 2 hours with primary antibody (diluted to working concentration in PBS), washed twice in PBST, stained with the secondary antibody (diluted to working concentration in PBS containing 1:10000 diluted Hoechst chromatin stain, Invitrogen) for 45 minutes, and washed a final three times in PBST.

293T cells were prepared on coverslips and handled in a 24-well plate. The cells were fixed and permeabilised using cold (-20˚C) methanol for at least 20 minutes. Cells were then carefully washed three times in PBST, blocked in blocking solution for 45 minutes before incubating with primary antibody for 2h (diluted to working concentration in PBS). Cells were then washed three times in 1X PBS and stained with the secondary antibody (diluted to working concentration in PBS containing 1:10000 diluted Hoechst chromatin stain, Invitrogen) for 1h. Samples were then washed three times with 1X PBS and mounted onto microscope slides (VWR

International) in Mowiol (Calbiochem). The slides were then covered, left overnight at room temperature to dry and stored at room temperature in the dark.

Images were acquired on a Nikon Ti-Eclipse widefield inverted microscope (Nikon, 10X dry objective lens for iPSCs, 60X oil objective lens for 293Ts) equipped with the NIS-Elements software (Nikon). Excitation and emission filters specific for the secondary antibody and DAPI were used. Multiple fields of view were selected at various XY plate coordinates and at least 6 fields per condition were analysed. Selected .TIF files generated directly using the NIS-Elements software or the NIS Elements viewer software (Nikon) were exported and assembled in Adobe Photoshop to generate overlay images. Information on all the primary and secondary antibodies that were used to generate the data presented in this thesis can be found in Table 3, Annex 1, p. 154.

3.7.2 Transmission electron microscopy

Transmission EM microscopy was performed in collaboration with the Centre for Ultrastructural Imaging (CUI) of King's College. Samples were prepared by a technique called negative staining. The purified virus was applied on an EM grid and exposed to a heavy metal (uranyl) stain. The grids were washed and imaged on a transmission electron microscope FEI Tecnai T20 and the associated software.

3.8 AAV LIFE CYCLE AND REP FUNCTIONAL ASSAYS

3.8.1 Infectious particles production assay

This assay was developed during this thesis and was designed to test rapidly and simply the effects of Rep mutations on the overall AAV lifecycle. For the purpose of the studies presented here it was only used to test Rep mutants, but the same assay can be used to test mutations in any of the AAV proteins. 293T cells were triple-transfected using PEI (see section 3.2.4) with plasmids pTRUF11, mini-pDG (or Rep mutants variants of mini-pDG) and pHGTI-Adeno1 to produce rAAV2-GFP. After 72h, the cell supernatant was harvested and increasing volumes (1µl, 10µl and 100µl) of supernatant were used to transduce HeLa cells (as explained in section 3.3.5). The percentage of GFP-positive Hela cells was determined 48h post-infection by flow cytometry (see section 3.6.4).

3.8.2 Replication assays

Three methods were used for the assessment of AAV replication. In all cases, 293T cells were first transfected using PEI (see section 3.2.4), 6h later infected with adenovirus at MOI 5 to stimulate AAV replication or mock infected for no replication controls, and harvested 72h post-transfection.

The first method used to detect AAV replication was slot blot (section 3.5.2). The total AAV DNA present in the cell was detected using the Rep-specific radiolabelled probe and normalised to the input plasmid DNA detected using the Amp-specific probe (probes prepared as described in section 3.5.1).

The second method used was based on absolute quantification by qPCR (section 3.5.4). The total DNA was extracted from the cells (section 3.4.1) and serial dilutions were quantified by qPCR. Cap primers were used for quantification of the total AAV DNA that was normalised to the housekeeping gene control cyclophillin.

For the third method, small-molecular weight DNA was extracted using the Hirt extraction method (described in section 3.4.2). ~2µg of DNA per sample were digested with the DpnI RE to degrade the input plasmid DNA. 1µg of DNA was loaded on a Southern blot gel, migrated and transferred to a nitrocellulose membrane as described in section 3.5.4. The membranes obtained were probed and imaged using a Rep or Amp radiolabelled probe (see section 3.5.1) to reveal AAV DNA and input DNA, respectively. Success of the DpnI digestion is controlled by the Amp-probed membrane: no bands are visible if the DpnI digestion is successful, while AAV replication is apparent by the appearance of bands on the Rep-probed membrane (because of the DpnI digestion only newly replicated AAV DNA will be visible). In case of normal AAV replication, a band for the monomeric AAV genome at 4.7kb and a band for the dimeric replicative form of the AAV genome at 9.4kb are expected. Additional larger bands (larger concatamers) can also be visible. For non-digested DpnI samples, in addition to the replication intermediates described above, plasmid DNA should be detected in the same way in both Amp- and Rep-probed membranes.

3.8.3 Supercoiled (sc) DNA nicking assay

scDNA nicking activity of purified Rep68* and mutants (kindly provided by Dr Carlos Escalante) was assayed as described previously (Lamartina et al. 2000a; Petri et al. 2015). Briefly, assays were performed in 30µl reactions containing 30mM Hepes-KOH (pH 7.5), 7mM $MgCl_2$, 0.5mM DTT, 4mM ATP, 40mM creatine phostate (Sigma), 1µg creatine phosphokinase (Sigma), in 15mM NaCl final concentration. 100ng scDNA plasmid and 200ng of purified Rep68* were added to the reactions. To test Rep68* scDNA nicking activity, different amounts of purified protein and 100ng of scDNA plasmid were incubated in a buffer containing 30mM Hepes-KOH (pH 7.5), 7mM $MgCl_2$ and 0.5mM DTT, in 50nM NaCl. All samples were incubated for 1h at 37°C and terminated by adding 10µl of stop reaction (proteinase K [1.2 µg/µl], 0.5% SDS and 30mM EDTA pH7.5) and incubating for 1h at 37°C. Samples were resolved in 1% agarose gel (1X TAE), which was subsequently stained with ethidium bromide (0.3µg/ml) in 1X TAE. scDNA plasmids used in this assay were pRVK (contains *AAVS1* nucleotides 1 to 3536 including the RBS and TRS sites) and pRVK-mutTRS (contains a mutated TRS that is not recognised by Rep) amplified in TOP10 competent cells (Petri et al. 2015).

3.9 iPSCs GENERATION AND CULTURE

3.9.1 Generation of iPSCs from dermal fibroblasts

Patient-derived human dermal fibroblasts at passage 6 were plated in fibroblast media (IMDM, GIBCO, with 10% FBS) onto gelatin-coated 6-well plates. For mycoplasma prevention, Plasmocin™ (Invivogen) was used at a prophylactic concentration of 2.5µg/ml in all culture media. Cells were then transduced at various MOIs with an excisable lentiviral vector carrying the four reprogramming factors Oct4, Sox2, Klf4, and c-Myc (Somers et al. 2010). After 24h, cell media was replaced with fresh iPSC media. iPSC medium is IMDM containing 10% FBS, 1nM of non-animal L-glutamine (Invitrogen), 1X sodium pyruvate (Invitrogen), 1X non-essential AA (NEAA, Invitrogen), 1X antibiotics, 0.1mM beta-mercaptoethanol (Invitrogen), 4ng/bFGF (from a 100ug/ml stock, Miltenyi Biotec) and 50ng/ml ascorbic acid (Sigma). At day 5, the cells were passaged onto 10cm dishes

containing irradiated mouse embryonic fibroblasts (iMEFs), and at day 10 the culture medium was switched to hESC medium containing DMEM-F12 (GIBCO), 20% Knockout serum replacement (GIBCO), 1mM L-glutamine, 0.1mM beta-mercaptoethanol, 1% NEAA, 1X antibiotics, 1X Na pyruvate, 7.5% NA bicarbonate, and 10ng/ml of bFGF. At this point, the medium was changed every 2-3 days and cells were left in culture for at least 28 days. Subsequently, the colonies with the most promising morphologies were hand picked and placed in single wells of a 24-well plate. At this point, medium was changed daily and cells were passaged when necessary using TrypLE (GIBCO).

3.9.2 High-sensitivity mycoplasma detection

The LookOut Mycoplasma PCR Detection Kit (Sigma) was used for high-sensitivity mycoplasma detection as per manufacturer's instructions, using the suggested JumpStart *Taq* DNA polymerase (Sigma). Briefly, supernatants from cells pre-cultured in the absence of anti-mycoplasma agents for several days were boiled for 5 minutes and centrifuged to pellet cellular debris. PCRs were performed in 25µl combining 2µl of cellular supernatant and 23µl of DNA polymerase/Rehydration Buffer mix (provided). Cycling conditions were as follows: 2 minutes at 94˚C, 40 cycles of 30 seconds at 94˚C followed by 30 seconds at 55˚C and 40 seconds at 42˚C, and a final cool down step to 4˚C. 8µl for each PCR sample were loaded on a 1.2% agarose gel and imaged as described in section 3.1.4. Negative samples show a single band at 481bp, while positive samples show a band at 260bp.

3.9.3 Mycoplasma treatment

In an attempt to eradicate mycoplasma contamination of iPSCs, Plasmocin^{TM} (Invivogen) was used at a curative concentration of 25 µg/ml in complete hES or fibroblast medium for 2 to 3 weeks.

CHAPTER 4: STRUCTURAL DETERMINANTS OF REP68 OLIGOMERISATION

The data presented in this chapter is part of two publications (Zarate-Perez et al. 2012; Zarate-Perez et al. 2013) (see also Annex 2, p. 162) that were generated in close collaboration with the laboratory of Dr Carlos Escalante at Virginia Commonwealth University (Richmond, Virginia, USA). My role in these publications was to investigate the functional consequences of mutations introduced into the Rep78/68 proteins. The structural findings presented in this chapter, including the data on the oligomerisation of the large Rep proteins and the *in vitro* functional assays were generated in Dr Escalante's lab. Where necessary to ensure that the implications of our findings are clear to the reader, contributions from Dr Escalante's lab are included and clearly indicated in the figure legends.

4.1 INTRODUCTION

The genomes of small DNA viruses encode for specialised initiator proteins that are essential for viral DNA replication. The Rep proteins from AAV, the LTag protein from SV40 and E1 protein from PV are amongst the best-characterised examples. These proteins belong to the SF3 of helicases, structurally defined by the presence of an AAA+ motor domain (Hickman and Dyda 2005). The combination of an AAA+ motor domain with an origin-binding domain confers a striking versatility to these proteins, allowing them to act as initiator proteins rather than mere helicases. In the case of AAV, Rep78/68 recognise the origin of replication (*ori*) through the RBS, and then unwind and nick the TRS to start a new cycle of DNA replication (Im and Muzyczka 1990; Owens et al. 1993; Brister and Muzyczka 2000). Similarly, SV40-LTag and PV-E1 bind to their respective *ori* were they promote melting of the origin sequences and initiation of replication (Fanning and Zhao 2009; Bergvall et al. 2013). Unwinding by SF3 helicases is ATP-dependent and requires the formation of an active helicase complex, which, in the case of SV40-LTag and PV-E1, is a double-hexameric ring (Fouts et al. 1999; Valle et al. 2000). Formation of this complex may be induced by the presence of DNA and/or ATP. The oligomerisation behaviour of

AAV Rep and its functional importance in the AAV life cycle, however, are less clear. Importantly, there are relevant functional differences between Rep and SV40-LTag and PV-E1. Notably, the OBD of Rep contains an endonuclease function that is required for DNA replication and Rep-mediated genomic integration. This function is absent in SV40-LTag and PV-E1 despite structural similarities, due to the loss of the HUH and Y motifs that form the endonuclease active site.

The first indications that the large Rep proteins could oligomerise came from several studies that reported the formation of multiple species in gel mobility shift assay experiments used to study Rep78/68 binding to RBS-containing DNA (McCarty et al. 1994b; Kyostio et al. 1995; Weitzman et al. 1996; Smith et al. 1997). Investigations stemming from these observations described Rep-Rep interactions for Rep78 and Rep68, and it was suggested that the presence of RBS-containing DNA and ATP could promote the formation of multimers (Hermonat and Batchu 1997; Di Pasquale and Stacey 1998; Davis et al. 1999). Moreover, using Rep68 and C-terminally truncated Rep mutants that retained endonuclease activity and RBS-binding, Weitzman et al. showed that Rep proteins could form hetero-oligomers on AAV hairpin DNA (Weitzman et al. 1996). Because of the structural similarities of AAV Rep with SV40 LTag and PV E1, it is reasonable to assume that the large Rep proteins also form hexameric rings. In support of this hypothesis, two studies have suggested that Rep forms hexamers in the presence of *ori* DNA (Smith et al. 1997; Dignam et al. 2007b). Smith and colleagues were the first to investigate the stoichiometry of Rep-DNA complexes. Using gel chromatography and crosslinking experiments, the authors showed that in the presence of a dsDNA substrate Rep78 formed a complex suggestive of a hexamer. The formation of the Rep78 complex was dependant on the presence of DNA but not ATP; in the absence of DNA Rep78 remained monomeric (Smith et al. 1997). Dignam et al. later reported that activation of the ATPase activity of Rep required the formation of Rep-DNA complexes and was stimulated by the presence of *ori* sequences and, to a lower extent, non-specific DNA. Rep78 and Rep68 showed a concentration-dependent oligomeric behaviour, both in the presence and absence of DNA, and the oligomeric species observed by gel chromatography were consistent with a hexamer bound to two copies of RBS-containing DNA (Dignam et al. 2007b). Li and colleagues had

previously reported a similar concentration-dependent effect on oligomerisation that also suggested that more than one copy of DNA was bound per complex, but the authors did not investigate the number of Rep68 molecules bound to the DNA (Li et al. 2003b). The observation that Rep68 complexes could bind more than one copy of DNA is also consistent with the intriguing report that Rep complexes can bind simultaneously to the ITR and *AAVS1* sequences (Weitzman et al. 1994).

Other reports, however, have described the formation of complexes other than hexamers. The crystal structure of the AAV5 Rep OBD bound to RBS-containing DNA showed five OBDs bound to one DNA molecule (Hickman et al. 2004). However, when a longer DNA substrate and/or full-length Rep68 were used, formation of larger complexes was observed (Hickman et al. 2004). A more recent study by Mansilla-Soto and colleagues also described a pentameric complex in the presence of a short RBS-containing dsDNA, and the authors suggest that this might be an intermediate assembly step towards a larger Rep68-ITR complex (Mansilla-Soto et al. 2009). This report went further to demonstrate how different DNA substrates can modulate the formation of distinct oligomeric species (Mansilla-Soto et al. 2009). Surprisingly, the authors found that in the presence of an unspecific ssDNA or heteroduplex (ssDNA-dsDNA) substrate Rep68 forms a large double-octameric complex. This complex was purified and shown to unwind the heteroduplex substrate in the presence of ATP and magnesium (Mansilla-Soto et al. 2009). The two octameric rings in this complex are oriented head-to-head, suggestive of a bidirectional activity. However, the current AAV replication model, unlike that of SV40 and PV, doesn't predict a bidirectional mechanism. SV40 and PV have a dsDNA genome, and their respective *ori* contain inverted repeats that are recognised by the initiator proteins and promote the formation of a head-to-head double hexameric complex necessary for bi-directional replication (Fanning and Zhao 2009; Bergvall et al. 2013). The AAV ITRs, however, only contain a series of direct repeats. Thus, it is not clear how a bidirectional Rep complex could function, although a few plausible scenarios have been suggested (Mansilla-Soto et al. 2009). First, if we assume that Rep complexes can bind more than one DNA molecule (Weitzman et al. 1994; Li et al. 2003b; Dignam et al. 2007b), a bidirectional complex could mediate the synchronised resolution (step e in Figure 3) of two separate ITRs.

Alternatively, the double-octameric complex could coordinate ITR refolding (after terminal resolution and duplication, step g in Figure 3, p. 26) and unwinding of double-stranded AAV DNA to promote further rounds of replication.

The lack of full-length Rep68 or Rep78 structures complicates the study of the determinants of oligomerisation as well as the characterisation of functionally relevant oligomeric complexes. Despite the fact that separate structures of the AAV5 Rep OBD and AAA+ helicase domain have been solved, it is not known how these two domains are oriented with respect to one another in the full-length Rep78/68. Because the large Rep proteins have a tendency to aggregate at high concentrations and low ionic strength conditions, it is difficult to study their oligomeric behaviour in the absence of DNA let alone the mechanisms of complex assembly (Smith et al. 1997; Dignam et al. 2007b; Mansilla-Soto et al. 2009). In addition, the tendency of the large Rep proteins to aggregate hinders structural studies of full-length Rep68 and Rep78. In a systematic analysis of Rep68 aggregation, our collaborators determined that only reducing agents decreased the aggregation significantly (Zarate-Perez et al. 2013). This implies that aggregation is induced by oxidation, potentially through the formation of intermolecular disulphide bonds, as was previously suggested by Smith and colleagues (Smith et al. 1997). This led us to investigating the roles of specific cysteine residues in Rep78/68 aggregation, and whether these residues are functionally important for the AAV life cycle.

In the context of AAV Rep oligomerisation, it is important to note that Rep40 and Rep52 remain monomeric in solution (Smith and Kotin 1998; James et al. 2003; James et al. 2004; Mansilla-Soto et al. 2009). This is in contrast to the behaviour of the corresponding domains of LTag and E1, which readily form hexamers (Sedman and Stenlund 1998; Enemark and Joshua-Tor 2006). Importantly, it has been shown that the Rep OBD is also monomeric in the absence of DNA, implying that both domains cooperate in the formation of oligomers (Mansilla-Soto et al. 2009). Interestingly, a previous study using small deletion mutants of Rep68 in an attempt to identify regions important for Rep68-Rep68 interactions indicated residues 151-188 (a putative coiled-coiled region) and 334-347 (part of the AAA+ active site) as important for multimerisation, supporting the notion that both domains participate

to oligomeric interfaces (Smith et al. 1997). A careful comparison of the oligomerisation interfaces of SV40-LTag and PV-E1 reveals an important similarity: both the core AAA+ and the α-helical bundle that connects to the origin-binding domain, contribute to the formation of a large hydrophobic oligomerisation interface (Figure 10B).

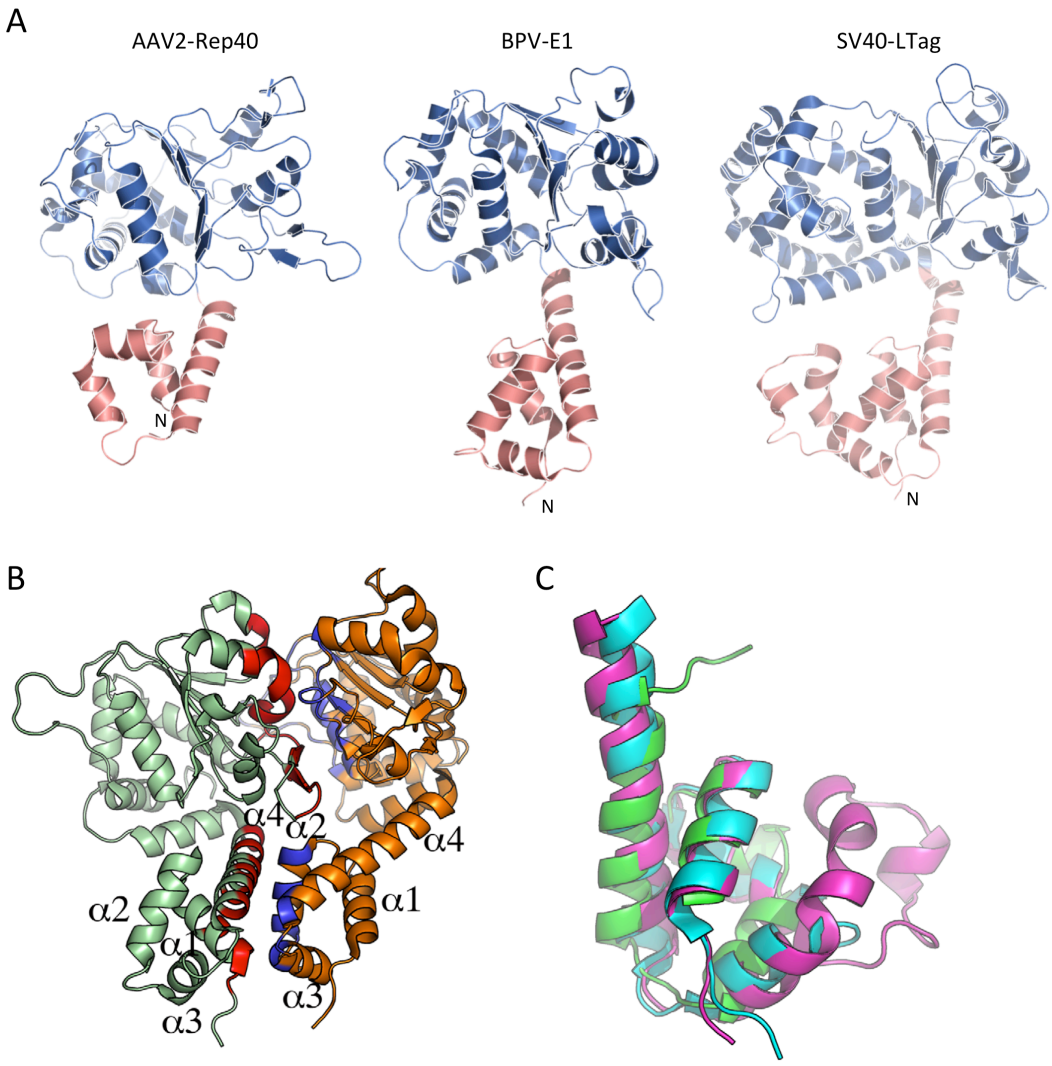


Figure 10: Structural comparison of SF3 helicases.
(A) Ribbon representations of AAV2-Rep40, PV-E1 and SV40-LTag. The core AAA+-like fold and the small helical bundle defined as OD in PV-E1 and SV40-LTag are represented in blue and salmon colours, respectively. **(B)** Dimer of PV-E1 with residues participating in the hydrophobic interface highlighted in blue and red. **(C)** Structural alignment of the helical bundles of AAV2-Rep40 (green), PV-E1 (cyan) and SV40-LTag (magenta). Figure adapted from (Zarate-Perez et al. 2012). Panels B and C are courtesy of C. Escalante.

Importantly, one side of the AAA+ part of the large oligomerisation interface includes all the catalytic residues of the AAA+ domain. Residues in the small α-helical bundle, termed the oligomerisation domain (OD) in PV-E1, are also part of the interface. In particular, two α-helices in the OD are involved in the formation of the hydrophobic interface (Figure 10B). In Rep40, this potential oligomerisation domain is significantly shorter than in SV40-LTag and PV-E1, measuring only 52 residues compared to 89 AAs and 68 AAs in SV40-LTag and PV-E1, respectively (Figure 10A and C). The consequent decrease in accessible surface could be sufficient to explain the lack of oligomerisation of Rep52/40.

In this chapter, I will present our contribution to the advancement in understanding of determinants Rep68 oligomerisation. Following the identification of cysteine disulphide bond formation as the main cause of Rep68 aggregation *in vitro*, we investigated the role of two exposed cysteine residues in the formation of aggregates. We identified a Rep68 cysteine mutant that prevents Rep68 aggregation even at high concentrations. In addition, our results demonstrate that this mutant is fully able to support the AAV lifecycle, and thus is a biologically relevant and valuable tool for structural studies of AAV Rep68. This mutant was then used to characterise the concentration-dependent dynamic oligomeric behaviour of Rep68. Subsequently, we investigated whether the shorter OD in Rep52/40 compared to SV40-LTag and PV-E1 could be the reason behind the lack of oligomerisation displayed by the small Rep proteins. We demonstrate that the interdomain linker of Rep78/68 plays a crucial role in Rep68 oligomerisation, and that its presence is sufficient to initiate oligomerisation of both the OBD and the helicase domains of Rep68. Furthermore, we identified an aromatic residue in the linker that is conserved in SF3 helicases and that is critical for Rep78/68 functions in support of the AAV life cycle.

4.2 RESULTS

4.2.1 Mutation of cysteine 151 to serine prevents aggregation.

To understand the role of cysteines in Rep68 aggregation, we analysed the crystal structure of Rep40 (James et al. 2003) and an AAV2 OBD model based on the

AAV5 OBD crystal structure (Hickman et al. 2002). There are six cysteine residues in Rep68: four are buried and semi-buried in the protein while two, C151 and C405, are exposed to the solvent (Figure 11A). C151 is part of the OBD and is positioned right next to tyrosine 152, which is part of the Y motif of Rep68 (see section 1.2.2). Despite the fact that Y152 is not the active site tyrosine that is directly involved in the catalysis of ssDNA, it has nonetheless been suggested to be important for the endonuclease activity of Rep68 (Davis et al. 2000). C405, on the other hand, is located in the helicase domain, more specifically within the pre-sensor 1 β-hairpin that is implicated in DNA translocation during DNA unwinding (Yoon-Robarts et al. 2004). Consequences of mutation C405 have never been studied, but mutants K404A and K406A have been shown to disrupt the helicase activity of Rep40 and Rep68. Interestingly, C405 is conserved in other AAV serotypes, while C151 is not. In the corresponding position in AAV5 a serine residue can be found instead of the cysteine (Figure 11B). Analytical gel filtration profiles of Rep68 containing the C151S mutation or the C405S mutation were compared to the profile of Rep68 WT. Aggregates of Rep68 elute with the void volume, suggesting their size is larger than the exclusion limit of the column used for this study. Mutation of C151 to serine (C151S) prevents the formation of aggregates, while the C405S mutant only reduces Rep68 aggregation (Figure 11C). Thus, cysteine 151 in the OBD but not cysteine 405 in the helicase domain seems to be critical in the formation of the intermolecular disulphide bonds that cause Rep68 aggregation in solution.

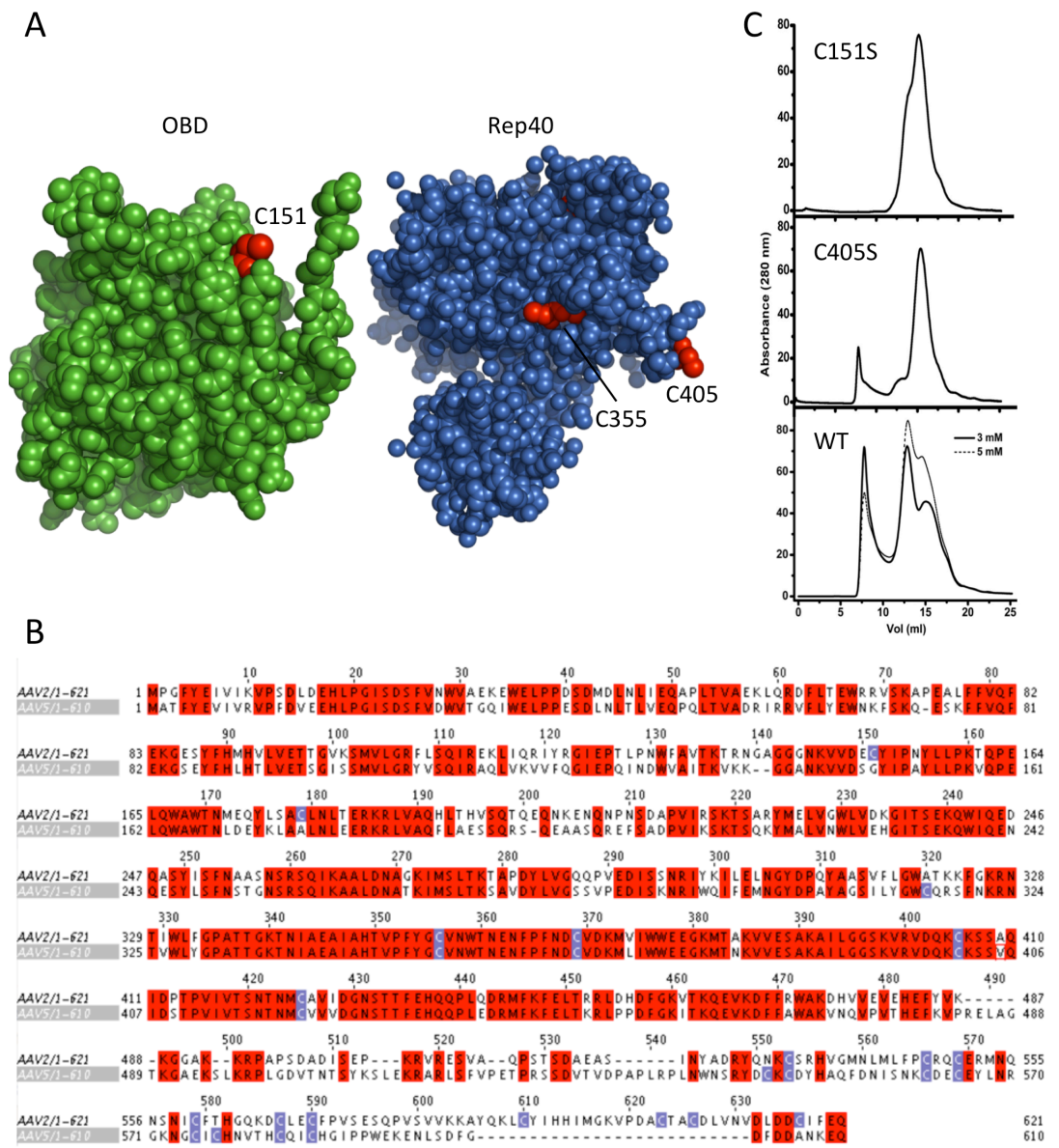


Figure 11: Mutation of cysteine residue C151 to serine prevents aggregation of Rep68.
(A) Sphere representation of a model of AAV2 OBD and of Rep40. Cysteine residues are highlighted in red. The two cysteine residues that appear to be fully exposed to the solvent, C151 and C405, are highlighted in red. The semi-buried cysteine C355 is also indicated. **(B)** Sequence alignment of AAV2 and AAV5. Identical residues are highlighted in red, cysteine residues are shown in purple. The alignment was performed with the ClustalW2 programme. **(C)** Analytical gel filtration profiles of Rep68 WT, Rep68-C151S and Rep68-C405S. The oligomerisation of WT Rep68 was assessed at two concentrations of the reducing agent TCEP. Panel C is courtesy of C. Escalante.

4.2.2 Rep68-C151S is functionally equivalent to Rep68 WT.

To be able to use Rep68-C151S for further structural and biochemical studies, we first needed to determine if this mutant was functional. Although C151 has never been identified in mutagenesis screens aimed at identifying residues important for

the catalytic activity of the large Rep proteins and serine is the most conservative AA substitution for cysteine, the C151S mutation could have detrimental effects on the functions of Rep68 due to its vicinity to the Y motif. We initially compared the biochemical activities of Rep68-C151S to those of Rep68 WT using *in vitro* functional assays. ATPase activity was assessed by a colorimetric assay that measures the amount of free phosphate in a reaction (Figure 12A). Helicase activity was determined by measuring DNA unwinding in a fluorescence resonance energy transfer (FRET)-based fluorimetric assay (Figure 12B), while DNA binding was assessed by fluorescence anisotropy (Figure 12C). Rep68-C151S performed as well as WT Rep68 in all these *in vitro* assays (Figure 12A, B and C). Then, we tested the ability of the mutant Rep to support the AAV life cycle. First we assessed if Rep68-C151S could efficiently replicate AAV DNA in 293T cells transfected with the pAV2 plasmid (WT or containing the Rep78/68-C151S mutant) and infected with adenovirus to stimulate AAV replication. The amount of AAV DNA present in the cells was quantified by slot blot using a Rep specific radiolabelled probe. Rep78/68 containing the C151S mutation proved to replicate AAV DNA as efficiently as WT Rep78/68 in the presence of adenovirus (Figure 12D). To determine if the Rep78/68-C151S mutant was also able to support the production of infectious viral particles, we infected HeLa cells using increasing volumes of supernatant from 293T cells producing recombinant AAV2-GFP in the context of the mutant Rep protein. Figure 12E shows that infectious rAAV2-GFP particles were efficiently formed in the presence of Rep78/68-C151S, albeit at slightly lower titres than WT Rep. Because the Rep78/68 supported AAV replication to WT levels, the cause of this small difference in infectious particles production is likely to be downstream of the DNA replication. To assess if defects in the efficiency of DNA packaging were the cause of this difference, we produced and purified WT and Rep-C151S AAV2 and assessed the fraction of full particles in our preparations by transmission electron microscopy. A significantly lower percentage of full particles is an indication of inefficient packaging of the viral DNA. The presence of the mutation in the Rep-C151S AAV virus was confirmed by sequencing (data not shown), but it did not affect the ratio of empty to full particles (Figure 12F).

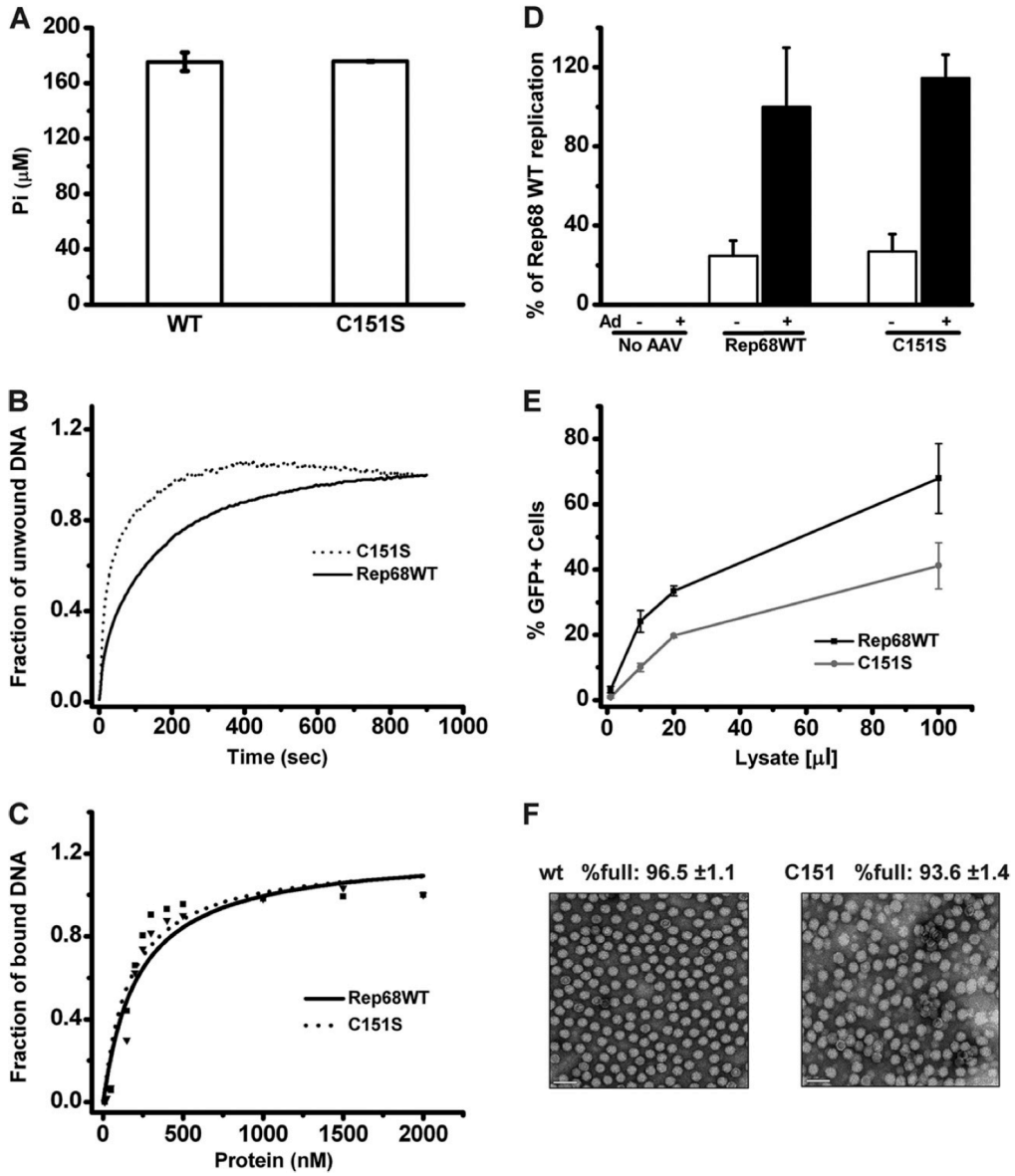


Figure 12: Functional comparison of Rep68 WT and Rep68-C151S.
(A) ATPase activity, **(B)** helicase activity and **(C)** DNA binding activities showed no significant difference between the WT and C151S proteins. **(D)** Determination of viral replication in the presence of WT or C151S mutant Rep. The bar graph shows the quantification of three slot blot experiments. The signal obtained with the Rep-specific probe was normalized to that obtained with the ampicillin-specific probe, to normalise for input (transfected) DNA. The replication of the WT AAV2 genome in the presence of adenovirus was set as 100%. Data are presented as mean ± s.e.m from three experiments. **(E)** Comparison of the production of rAAV2-GFP infectious particles in the presence of WT or C151S Rep. rAAV2-GFP particles were produced in 293T cells in the presence of WT or C151S Rep. Various volumes of crude lysate (in µl, x axis) were added to HeLa cells, and the percentage of GFP-positive infected cells was determined by FACS analysis. Data from four experiments are presented as mean ± s.e.m. **(F)** Visualisation of WT AAV2 and AAV2-RepC151S viral particles by transmission electron microscopy at 50,000X magnification. Empty viral particles appear as white rings. For each sample, 6 fields of approximately 200 particles were counted; the mean percentage of full particles and the standard deviation are indicated. Panels **(A), (B)** and **(C)** are courtesy of C. Escalante.

Taken together, these results demonstrate that Rep78/68-C151S is functionally comparable to WT Rep78/68 in supporting AAV replication and, more generally, the AAV life cycle. Furthermore, the biochemical properties of Rep68, including RBS binding, DNA unwinding and the ATPase activity, are recapitulated by Rep68-C151S. We conclude that Rep68-C151S is sufficiently similar to Rep68 WT and thus that this mutant is biologically relevant and can be used for further structural and biochemical studies. We have used this mutant as the new standard for the characterisation of the oligomeric behaviour of Rep68, and I will refer to Rep68-C151S as Rep68* henceforth.

4.2.3 The interdomain linker is essential for Rep68 oligomerisation.

Identifying the structural features that promote oligomerisation is pivotal to understanding the mechanisms of assembly of functionally relevant Rep complexes. The structural comparison of SF3 helicases revealed that the monomeric small AAV Rep proteins have a shorter OD compared to that of the hexameric SV40-Tag and PV-E1 proteins (see Figure 10). In the large Rep proteins, Rep78 and Rep68, this shorter OD is connected to the OBD via a linker that spans residues 200-224. To investigate whether this linker extends the OD in Rep68/78 and contributes to oligomerisation, we carried out secondary structure prediction on the linker. The results suggest that while most of the linker is predicted to be disordered, residues 219 to 224 have the potential to form an α-helix, possibly extending the N-terminal α -helix of the OD and increasing the surface accessible area (Figure 13A). To test if this part of the linker functionally belongs to the OD, we assessed the oligomerisation properties of a truncated Rep missing the OBD but including the linker (Rep68Δ209) by sedimentation velocity experiments. These experiments, performed by analytical ultracentrifugation, allow the determination of size distribution functions, which provide information on the protein species present in a sample. From the known mass of a protein or complex it is possible to calculate a theoretical sedimentation coefficient, which can then be used to infer the identity of the species (peaks) observed in the size distribution functions. In this case, Figure 13B shows that in contrast to Rep40, which is monomeric, Rep68Δ209 can form

oligomers. Surprisingly, addition of the linker to the OBD also induced oligomerisation (Figure 13C). Thus, the linker is sufficient to induce oligomerisation of both domains of Rep68, which are otherwise monomeric. Taken together, these results provide evidence that the interdomain linker of Rep68 is a functional component of the OD and establish its critical role in Rep68 oligomerisation.

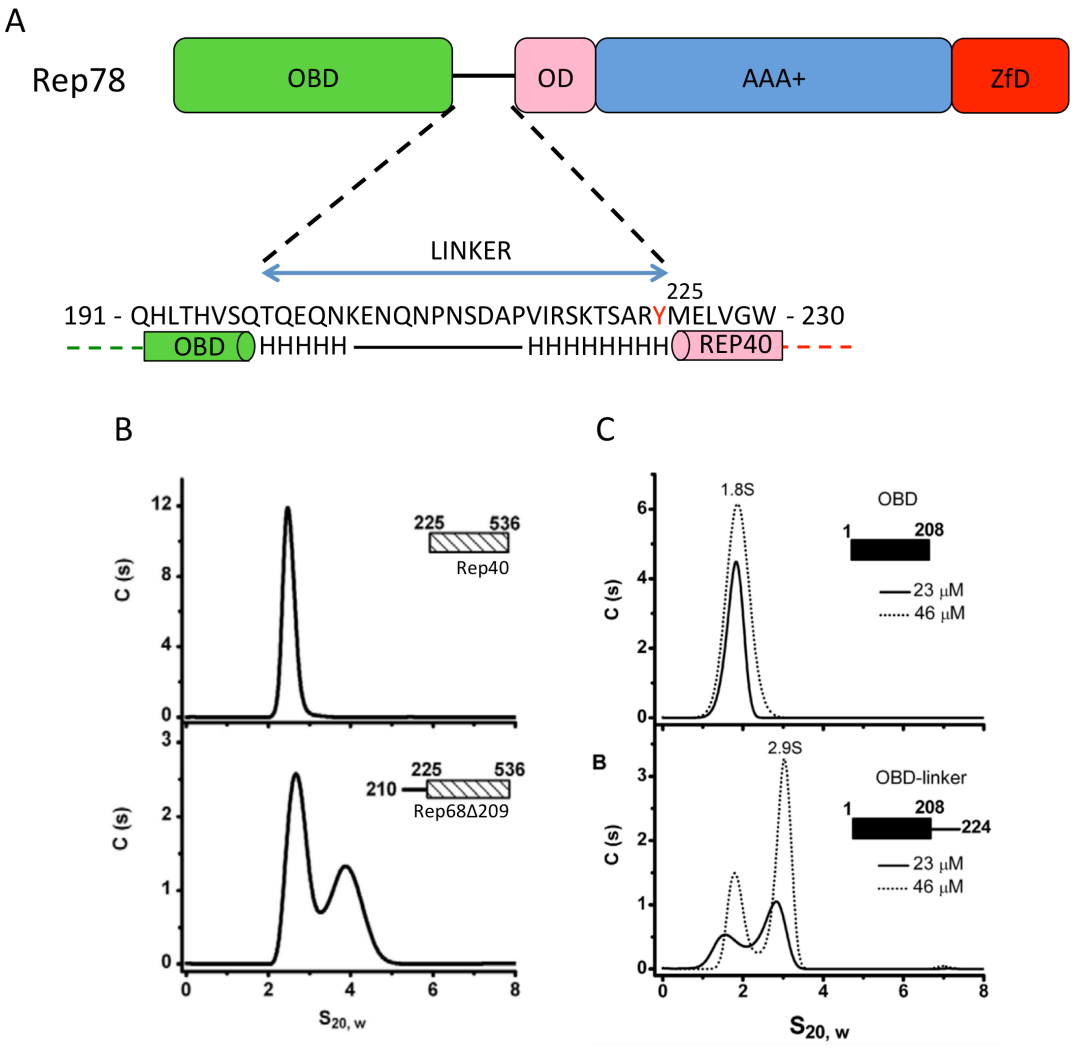


Figure 13: The interdomain linker of Rep68 is essential for oligomerisation.
(A) Schematic diagram of the domain architecture of Rep78, including the OBD (green), the OD (pink), the helicase domain (blue) and the ZnF domain (red). The sequence of the linker is shown below together with the predicted helical regions indicated by H. Sedimentation profiles of **(B)** Rep40 and Rep68Δ209 and of the OBD and the OBD-linker **(C)**. Panels **(B)** and **(C)** are courtesy of C. Escalante.

4.2.4 Linker residue Y224 is critical for Rep68 oligomerisation and function and represents a conserved feature in SF3 helicases

A structural model of the helicase domain of Rep68 with the addition of the last 7 residues from the linker (Rep68Δ217) was generated using the available Rep40 structure and the linker residues added as a helical extension. This model was then used for structural alignment with the OD domains of SV40-LTag and PV-E1. The alignment reveals that the modelled α-helix superimposes with helix 1 of SV40-LTag and PV-E1 helicase domains. Furthermore, Rep residue Y224 (the last residue of the linker) corresponds to aromatic residues W270 and F313 of SV40-LTag and PV-E1, respectively (Figure 14 A and B). These aromatic residues appear to be an integral part of the oligomerisation interface in LTag and E1 hexameric rings. Thus, we tested the hypothesis that Y224 plays an equivalent role in Rep by mutating it to alanine and assessing the consequences on Rep68* oligomerisation and more generally on the AAV life cycle. Rep68*-Y224A showed severe oligomerisation (Figure 14C). Moreover, to assess if this oligomerisation defect caused by the Y224A mutation has any consequences on Rep functions and thus on the AAV life cycle, supernatant collected from cells producing rAAV2-GFP in cells transfected with WT or mutated Rep78/68 was then used to infect HeLa cells. Figure 14D shows that the supernatant collected from cells transfected with the mutant helper virus did not contain any infectious rAAV2-GFP, as determined by FACS analysis of GFP positive cells. Thus, Rep68*-Y224 was completely unable to support the production of infectious AAV particles. Taken together, these results suggest that AAV Rep residue Y224 and the oligomerisation properties it confers to Rep78/68 play a crucial role in supporting the AAV life cycle.

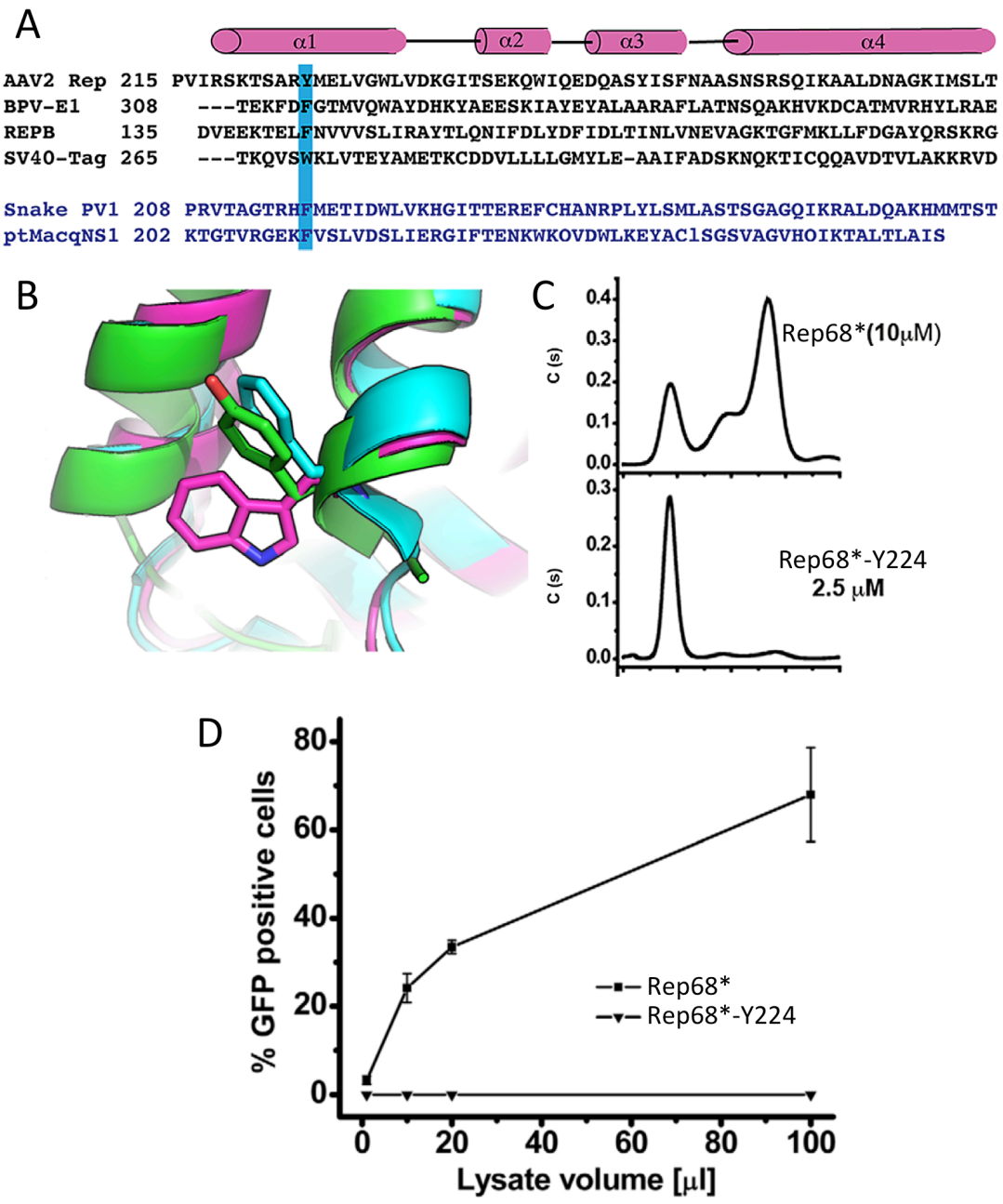


Figure 14: The conserved residue Y224 is essential for Rep68 oligomerisation and function.
(A) Structural alignment of the OD domains of AAV2-Rep, PV-E1, SV40-LTag and the bacterial RepB, and of two non-structural parvoviral proteins from snake parvovirus 1 and point-tailed macaque parvovirus NS1. A conserved aromatic residue is highlighted in blue. **(B)** Structure superposition of the conserved aromatic residue for AAV2-Rep (green), SV40-Tag (magenta) and PV-E1 (cyan). **(C)** Comparative analysis of the sedimentation profiles of Rep68 and Rep68-Y224A. **(D)** Consequences of the Y224A mutation on the production of rAAV2-GFP infectious particles in the presence of WT or Y224A Rep. Data from three experiments presented as mean ± s.e.m. Panels **(A), (B)** and **(C)** are courtesy of C. Escalante.

4.3 DISCUSSION

The role of SF3 helicases such as SV40-LTag or PV-E1 as initiators of DNA replication relies on their ability to oligomerise upon binding to the *ori* DNA. Similarly, it is thought that AAV Rep oligomerisation is also required for its functions during the AAV life cycle. However, while it is well established that SV40-LTag and PV-E1 form hexameric rings, the oligomeric behaviour of AAV Rep appears to be more flexible (Li et al. 2003b; Mansilla-Soto et al. 2009; Zarate-Perez et al. 2013). In particular, the small Rep proteins are monomeric, while the large Rep proteins form larger complexes. Pentameric, hexameric and octameric complexes have been suggested for Rep78/68 (Smith et al. 1997; Li et al. 2003b; Dignam et al. 2007b; Mansilla-Soto et al. 2009). However, studies of Rep oligomerisation and its further structural characterisation have been hindered by its tendency to aggregate (Smith et al. 1997; Dignam et al. 2007b; Mansilla-Soto et al. 2009). We have identified oxidation and the formation of intermolecular disulphide bonds as the cause of this aggregation. In particular, cysteine 151 seems to be involved in this process, and its mutation to serine prevents Rep68 aggregation. Intriguingly, while it has never been possible to obtain a crystal structure for Rep68 or the OBD of AAV2, the crystal structure of the Rep OBD from AAV5, which contains a serine in the corresponding position, was solved. Importantly, we have shown that Rep68 containing this conservative mutation is functional, rendering Rep68* (Rep68-C151S) a biologically relevant tool for the study of AAV2 Rep structure and oligomerisation. Rep68* was used for further characterisation of Rep oligomerisation in the absence of DNA, and revealed a complex and dynamic picture (Zarate-Perez et al. 2013). In particular, Rep68* showed a concentration and nucleotide dependent oligomerisation, forming a monomer-dimer equilibrium at low concentrations and heptameric and octameric rings at high concentrations (Figure 15). More recently, Rep68* proved useful in determining the crystal structure of the OBD and OBD-linker of AAV2 Rep, the first structural characterisation of the full-length Rep68, and the X-ray structure of the OBD of Rep68 bound to the *AAVS1* RBS-TRS site (Musayev et al. 2015a; Musayev et al. 2015b). The structure of the AAV2 OBD is very similar to the AAV5 structure that

was reported previously (Hickman et al. 2002), with the exception of the DNA binding loop and the C-terminal long α-helix (helix F), which are positioned at different angles. The linker was found to be partially structured, a property that was supported by small-angle X-ray scattering (SAXS) data of Rep68* that showed a compact molecule with all the functionally important residues positioned on one face of the protein (Musayev et al. 2015a). The study of Rep68* OBD binding to *AAVS1* identified a heptameric ring as the oligomer that binds *AAVS1* and the assembly of this complex required the OBD, the linker and the helicase domain. Furthermore, the X-ray structure of Rep68 bound to the *AAVS1* RBS-TRS site provides insights into the mechanism of *AAVS1* DNA recognition by Rep68* (Musayev et al. 2015b).

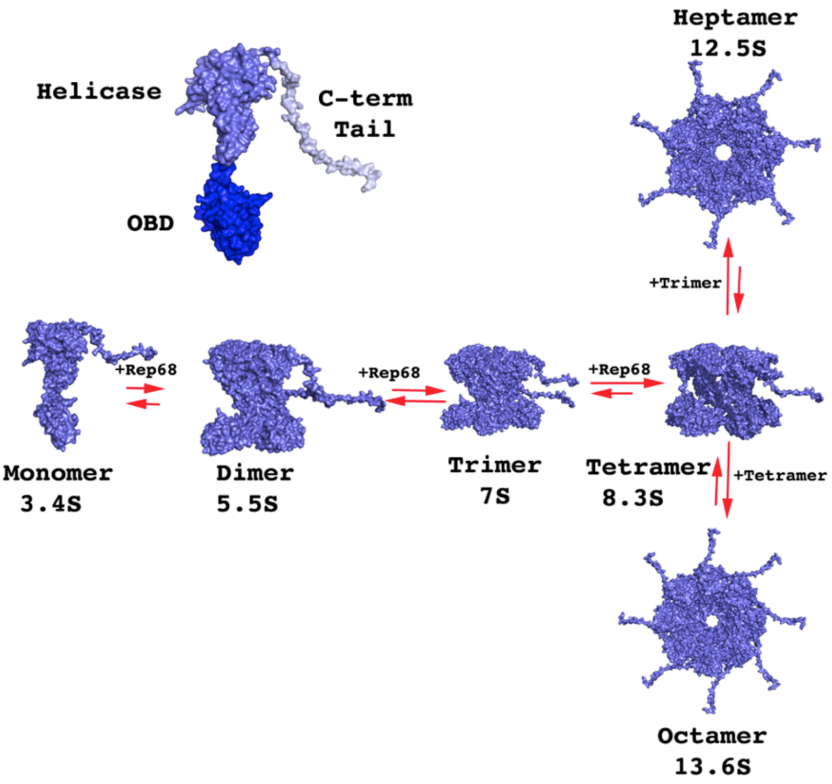


Figure 15: A model of the dynamic oligomerisation of Rep68.
A model of the dynamic oligomeric properties of Rep68*. A model for full-length Rep68 was built from available structures of the OBD (dark blue) and helicase domain (blue) and a modelled flexible C-terminal tail. The model is based on a systematic analysis of the oligomeric behaviour of Rep68* by analytical ultracentrifugation. Rep68* was shown to shift from a monomer-dimer equilibrium at low concentrations to the formation of large oligomeric rings, through intermediate trimeric and tetrameric complexes. Adapted from (Zarate-Perez et al. 2013).

Identifying the determinants of Rep oligomerisation is key to our understanding of the assembly of Rep complexes on the relevant DNA substrates and their mechanism of action. The analysis of the oligomeric interface formed by SV40-LTag and PV-E1 revealed a crucial role of the OD in stabilising the hexamers that are formed by these proteins (Titolo et al. 2000; Gai et al. 2004; Weisshart et al. 2004). In Rep40, the OD is significantly shorter and the available surface is not sufficient to stably oligomerise. We have shown that part of the interdomain linker of Rep is in fact an extension of the OD and promotes oligomerisation even in the absence of DNA. Secondary structure prediction shows that the C-terminal residues of the linker have the potential to extend the first helix of the SF3 helicase domain of Rep. Furthermore, substitution of the Rep linker with an unrelated and unstructured linker disrupts oligomerisation (Zarate-Perez et al. 2012). In agreement with our results, a recent report also highlighted the role of the linker in the oligomerisation of AAV5 Rep68 (Maggin et al. 2012). These results confirm the critical role of the OD in promoting and stabilising the oligomerisation of SF3 helicases. Importantly, constructs of SV40-LTag and PV-E1 helicase domains lacking the OD fail to oligomerise (Titolo et al. 2000; Gai et al. 2004; Weisshart et al. 2004). An OD that promotes dimerisation has also been mapped to the region that connects the OBD and the helicase domain of the non-structural protein (NS1) of MVM, another parvovirus. Perhaps not surprisingly, some proteins that belong to the HUH family of endonucleases, and thus are evolutionarily related to AAV Rep78/68 and NS1, also contain a functional oligomerisation domain. For example, the Rep proteins of plant geminiviruses, such as the tomato yellow leaf curl virus (TYLCV), share a domain architecture with AAV Rep, with an N-terminal HUH origin-binding domain, a central oligomerisation domain and a C-terminal SF3 helicase function (Fondong 2013). Another example is the RCR initiator RepB protein from the streptococcal plasmid pMV158, which assembles as a hexameric complex to initiate plasmid DNA replication (Boer et al. 2009). This protein lacks the C-terminal AAA+ motor domain, but contains a four-helical bundle OD similar to that of SF3 helicases located C-terminally to a HUH endonuclease domain.

AAV Rep residue Y224, the last AA before the start codon of the monomeric Rep52/40, plays a critical role in Rep68 oligomerisation. Strikingly, Rep containing

the Y224A mutation is unable to support the production of infectious AAV vectors. While this phenotype can be attributed to a defective oligomerisation of the Rep proteins, it is at present unclear which enzymatic activities are actually affected. Importantly, a more conservative mutation, Y224F, was previously identified as defective in ITR DNA binding as well as endonuclease, ATPase and helicase activities (Walker et al. 1997a). Because Y224 is not part of any of the enzymatic sites of Rep, these results suggest that Rep oligomerisation is important in all of these functions. The importance of this residue is also highlighted by the fact that its aromatic character appears to be conserved in non-structural proteins of other parvoviruses, in SF3 helicases, as well as in the prokaryotic RepB protein (Figure 14A).

It is remarkable that AAV appears to have evolved two sets of Rep proteins that are functional helicases by virtue of a shared helicase domain but differ in their ability to oligomerise. The small Rep proteins Rep52 and Rep40, generated from the p19 promoter, are monomeric due to their truncated OD, and only transiently form a dimer required for ATP hydrolysis (Dignam et al. 2007b; Zarate-Perez et al. 2012). Intriguingly, the last base pair before the start codon of Rep52/40 encodes for a conserved tyrosine residue that plays a crucial role in the oligomerisation of Rep78/68 and more generally in the AAV life cycle. Based on the oligomeric and functional differences between the large and the small Rep proteins, we speculate that AAV may utilise a single AAA+ helicase domain for two distinct mechanisms of unwinding. The mechanism of translocation by the transiently dimeric small Rep proteins could be more similar to that of dimeric helicases, implying either an inchworm mechanism or rolling mechanism of translocation (Patel and Donmez 2006). The large Rep proteins, on the other hand, may function as classic multimeric helicases like other SF3 helicases (Enemark and Joshua-Tor 2006). The process of DNA packaging mediated by the small Rep proteins, however, remains elusive, and it is plausible that multimerisation of these proteins is induced by interaction with the capsid proteins.

The results presented in this chapter describe how we have identified and addressed the causes of Rep68 aggregation that was hindering further structural studies of this remarkably multi-functional protein. Cysteine residue C151 was found to be critical for the formation of intermolecular disulphide bonds that cause

Rep68 aggregation, and its mutation to serine prevents the formation of aggregates. In addition, we demonstrated that the C151S mutant of Rep68 (or Rep68*) is functionally equivalent to Rep68 WT. Rep68* is proving a useful tool that can be used to improve our understanding of Rep68 structural biology. Apo-Rep68* was shown to have a dynamic and concentration-dependent oligomerisation, supporting previous reports that showed that Rep68 can form different oligomeric complexes in the presence of different DNA substrates (Li et al. 2003b; Mansilla-Soto et al. 2009). Furthermore, in our investigations of the structural determinants of Rep68 oligomerisation we identified a functional OD in Rep68 that includes part of the interdomain linker and is sufficient to initiate oligomerisation of the OBD and the helicase domain, which are otherwise monomeric. We described how the linker residue Y224, conserved in parvoviral non-structural proteins and SF3 helicases, plays a critical role in oligomerisation and, more generally, in the AAV life cycle. Further investigations of the oligomeric behaviour of Rep could thus provide insightful information into the mechanisms underlying the multiple functions of Rep during the AAV life cycle. The next step will be to investigate the functional importance of the oligomeric interfaces formed between Rep68 molecules and how these interfaces contribute to the dynamic oligomerisation of Rep68.

CHAPTER 5: IDENTIFICATION OF AN OLIGOMERIC INTERFACE ESSENTIAL FOR AAV REP FUNCTIONS

A manuscript based on the findings presented in this chapter is currently in preparation (see Annex 2, p. 162). This work combines structural findings from the laboratory of Dr Carlos Escalante with our thorough functional study on the consequences of disrupting potential oligomeric interfaces. Where necessary to ensure that the implications of our findings are clear to the reader, contributions from Dr Escalante's lab are included and clearly indicated in the figure legends.

5.1 INTRODUCTION

In the previous chapter, I have described how we addressed the problem of AAV Rep aggregation that was hindering further structural studies of these intriguing viral proteins. The discovery of a Rep mutant (Rep68*) that did not aggregate even at high concentrations and was fully functional allowed a meticulous characterisation of the oligomeric behaviour of apo-Rep68 (Zarate-Perez et al. 2013). The result is a complex and dynamic model of Rep68 oligomerisation, supporting the idea, initially introduced by Muzyczka and colleagues (Li et al. 2003b), that Rep78/68 can adopt different oligomeric states depending on the DNA substrate encountered, the presence of nucleotides, and the concentration of Rep78/68. The discovery that the interdomain linker of AAV Rep plays a crucial role in oligomerisation by providing part of the OD is an initial insight into the structural requirements for the functional oligomerisation of Rep proteins and provides a structural explanation for the observed lack of oligomerisation of the small Rep proteins. Moreover, it suggests important similarities between AAV Rep proteins and related SF3 helicases or HUH endonucleases proteins in terms of oligomerisation.

Identifying the protein-protein interfaces formed in specific oligomeric complexes would allow a better structural and mechanistic understanding of the enzymatic properties of the AAV Rep proteins. For example, SV40-LTag mutants that form hexamers but fail to form stable double hexamers have been shown to

perform well in a series of *in vitro* assays measuring DNA binding and unwinding, but fail to support viral replication, and have proved valuable in defining the bidirectional steps of SV40 replication (Weisshart et al. 1999). However very few studies have addressed the consequences of mutations in AAV Rep in the context of oligomerisation. These studies identified the C-terminal region of Rep, as well as short internal deletions of residues 151-188 and residues 334-347, as important for oligomerisation (Weitzman et al. 1996; Hermonat and Batchu 1997; Smith et al. 1997). The picture is complicated by the difficulty of distinguishing between defects in specific enzymatic functions and defects in oligomerisation that indirectly affect those functions. For this reason, it is possible that some studies on the functional consequences of Rep mutations have identified non-functional mutants defective for oligomerisation without knowing it. Mutants that fall in this category are likely to be Rep78/68 mutants that are defective in RBS binding, TRS nicking and helicase activity, such as the Y224F mutant (Walker et al. 1997b) or the P415H mutant (McCarty et al. 1992). Misfolded or degraded proteins, however, could also explain a completely inactive Rep78/68 mutant. Furthermore, mutations that affect DNA binding interfaces, such as the K136A and R138A Rep78 mutants identified by Urabe et al. (Urabe et al. 1999), could also affect oligomerisation, despite not being directly involved in the formation of oligomeric interfaces. Confirming this possibility, mutation of residue R107, which was initially identified as essential for DNA binding (Urabe et al. 1999), was later shown to also disrupt oligomerisation (Mansilla-Soto et al. 2009).

In this chapter, we take advantage of the possibility to use Rep68* to study the oligomeric properties of Rep in the absence of DNA to identify Rep oligomerisation mutants independently of DNA binding. Nevertheless, we anticipate that oligomerisation-deficient mutants will also fail to bind DNA. Furthermore, an additional goal of this study was to begin defining the oligomerisation requirements for specific enzymatic functions of Rep by discovering residues that are part of oligomeric interfaces but that do not affect the catalytic folds of the proteins.

5.2 RESULTS

5.2.1 Identification of potential oligomeric interfaces formed by AAV Rep

To identify potential functionally relevant Rep-Rep interfaces, we used two approaches. The first interface we decided to investigate is based on a Rep dimer modelled using the available structure of the C-terminal helicase domain extended as an α-helix to nucleotide 217 to mimic the PV-E1 dimer interface (Figure 16A). A closer analysis of this Rep-Rep interface reveals that the conserved aromatic residue Y224, which is part of the OD and plays a crucial role in oligomerisation (see Chapter 4), is an important component of this interface. Y224 is participating in the oligomeric interface by forming hydrophobic interactions with the isoleucine residue I251 and an hydrogen bond with the main chain carbonyl oxygen of Asn254 in the neighbouring Rep molecule (Figure 16B).

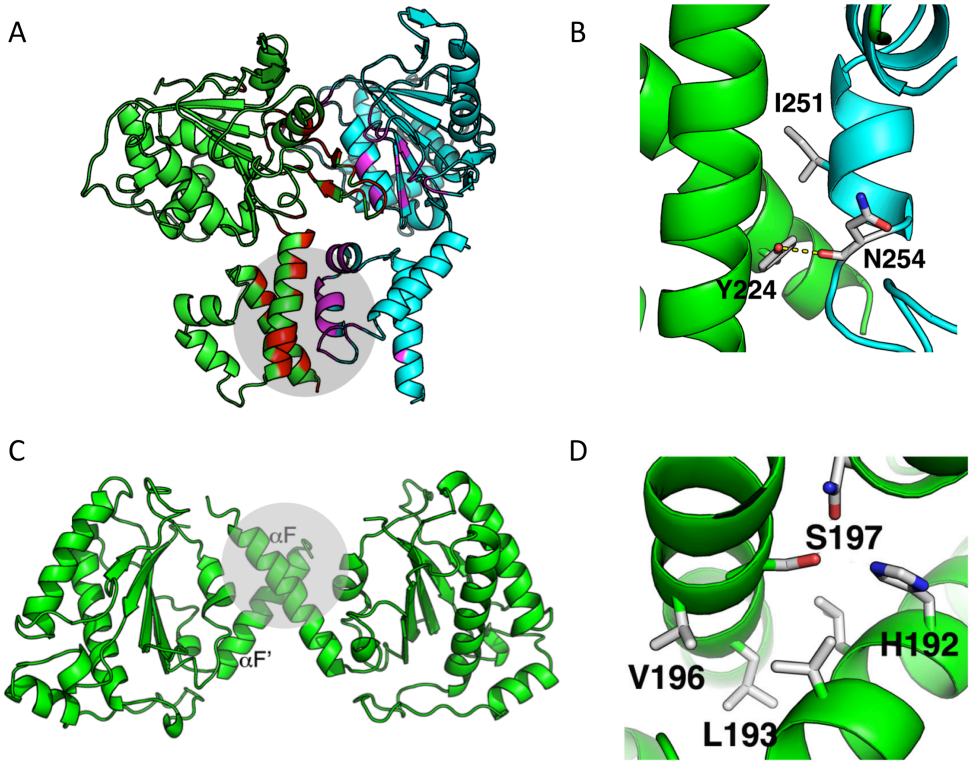


Figure 16: Identification of Rep-Rep interfaces.
(A) A Rep-Rep dimer was modelled based on the described PV-E1 dimer. One Rep monomer is depicted in green, the other in cyan. Residues that participate to the interface are highlighted in red and magenta, respectively. **(B)** Close-up on the interactions formed by the conserved aromatic residue Y224 within the modelled oligomeric interface. These interactions include the formation of a hydrogen bond with the main chain carbonyl oxygen of N254 and hydrophobic interactions with I251A. **(C)** The formation of an OBD-OBD dimer was reported in the crystallisation studies of AAV5 and AAV2 Rep OBD. **(D)** Close-up on the interface observed in crystallography studies. Residues L193 and V196 contribute to both sides of the hydrophobic interface. Figure courtesy of C. Escalante.

The second interface is the result of the crystallography studies of the Rep OBD from AAV2 (Musayev et al. 2015a). Two of three OBD molecules in the asymmetric unit consistently form a dimer through interactions between the long C-terminal α-helix (α-helix F), resembling a pseudo-coiled coil interaction (Figure 16C and D). A similar interface was also observed in structural studies of the AAV5 Rep OBD in complex with RBS DNA (Hickman et al. 2004). In AAV2, the residues involved in the formation of this interface include residues L193 and V196 making hydrophobic interactions and residues H192 and S197 forming a hydrogen bond (Figure 16D).

Based on these observations, we designed a series of mutants to test the relevance of these two interfaces for Rep functions during the AAV life cycle. While we have shown previously that the mutant Rep68-Y224A fails to oligomerise *in vitro* and that it cannot support the production of infectious AAV particles (see Chapter 3), we have not determined what the consequences of this mutation are on the diverse functions of Rep. In addition to Y224A, we have also tested the impact of the more conservative Y224F mutation. This mutation maintains the evolutionarily conserved aromatic character of this amino acid, but was shown previously to affect Rep functions, including DNA binding, nicking and unwinding (Walker et al. 1997a). We have also assessed the consequences of mutating the residue I251 (alone or in combination with Y224) to alanine. Finally, we have tested the double mutant L193A-V196A, which is part of the interface observed in crystallographic studies of Rep OBD. Importantly, all the mutants that we have selected are not part of the known enzymatic folds of Rep.

5.2.2 Consequences of mutations on the oligomerisation of Rep68*

We first studied the oligomeric behaviour of Rep68*-Y224A, Rep68*-Y224F, Rep68*-I251A, Rep68*-YI (Y224A-I251A) and Rep68*-LV (L193A-V196A) by analytical ultracentrifugation. Figure 17 shows the sedimentation profiles of Rep68* and the oligomerisation interface mutants. Confirming our previous observations (see Chapter 3), Rep68*-Y224A does not form the large complexes described for Rep68* (Zarate-Perez et al. 2013) and only forms small intermediate complexes. Rep68*-Y224F, on the other hand, seems to be able to form larger oligomeric complexes, although not to the same extent as Rep68*. Both the single

mutant Rep68*-I251A and the double mutant Rep68*-YI are mostly monomeric and fail to oligomerise under our experimental conditions. Finally, the oligomeric properties of the Rep68*-LV mutant, containing mutations in the interface observed in crystallography studies, are comparable to Rep68*.

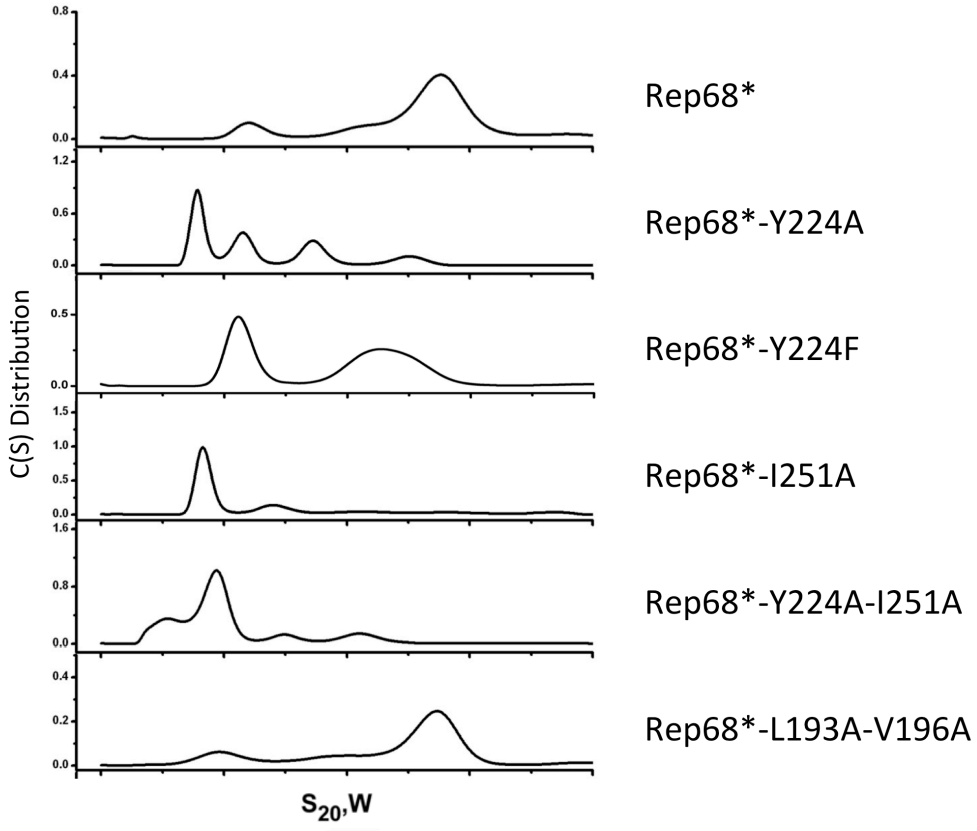


Figure 17: Sedimentation velocity analysis of the oligomeric properties of interface mutants.
The interface mutants identified in Figure 16 were assessed for their ability to oligomerise by analytical ultracentrifugation. Courtesy of C. Escalante.

5.2.3 Oligomerisation-deficient mutants fail to support the AAV life cycle

In view of their oligomerisation properties, we then sought to assess the ability of these mutants to support the AAV life cycle. First, we tested the performance of the oligomerisation mutants in producing infectious AAV-GFP particles. This assay allows for a very general assessment of the ability of Rep (and potentially Cap) mutants to support the production of AAV vectors (AAV2-GFP in this case as it provides an easy readout) and thus to support the viral life cycle. Defects in any step of the AAV life cycle, including viral DNA replication, packaging, and virus

release will be detected by a reduction in the amounts of infectious particles produced. However, the readout of this assay does not provide information on which stage of the life cycle is not supported by the mutants. As a negative control, we used the NTP-binding mutant K340H, which has been shown to be deficient for ATPase and helicase activities and does not support AAV replication, but is still able to oligomerise (Chejanovsky and Carter 1990; Owens et al. 1991; Kyostio and Owens 1996; Smith et al. 1997). Figure 18A shows that, not surprisingly, mutants defective in oligomerisation are also unable to produce infectious AAV particles. As described in Chapter 3, the Y224A Rep mutant does not support the production of infectious AAV-GFP particles. The monomeric I251A and Y224A-I251A mutants also fail to produce viral particles. The more conservative mutation Y224F, which retains some potential to oligomerise, is severely impaired but not entirely deficient in producing infectious AAV. The double mutant L193A-V196A, on the other hand, supports the production of infectious AAV-GFP to levels comparable to WT Rep and Rep68* (Figure 18A).

We then assessed the ability of the mutants to support AAV replication using two different assays. First, we quantified the total viral DNA present in the cells by qPCR (Figure 18B). The readout of this assay is the extent of AAV DNA amplification from the input plasmid DNA by Rep-mediated replication. We also isolated small molecular weight DNA by a modified Hirt extract procedure to study the replicative intermediates formed during AAV replication (Figure 18C). The interface mutants Y224A, I251A and the double mutant Y224A-I251A all failed to support AAV DNA replication in both assays. Similarly to what we observed in the infectious particle production assay, the Y224F mutant supported limited AAV replication in the presence of the helper adenovirus, while the L193A-V196A double mutant appears to efficiently replicate AAV DNA. In all cases where we observed DNA replication, the relative levels of monomeric and dimeric replicative intermediates were comparable (Figure 18C). The low level replication observed in the absence of adenovirus is due to the presence of the adenoviral E1A and E1B proteins in 293T cells (Graham et al. 1977). Thus, it appears that the defects we observed in the production of infectious viral particles (Figure 18A) are caused by the lack of replication of viral DNA.

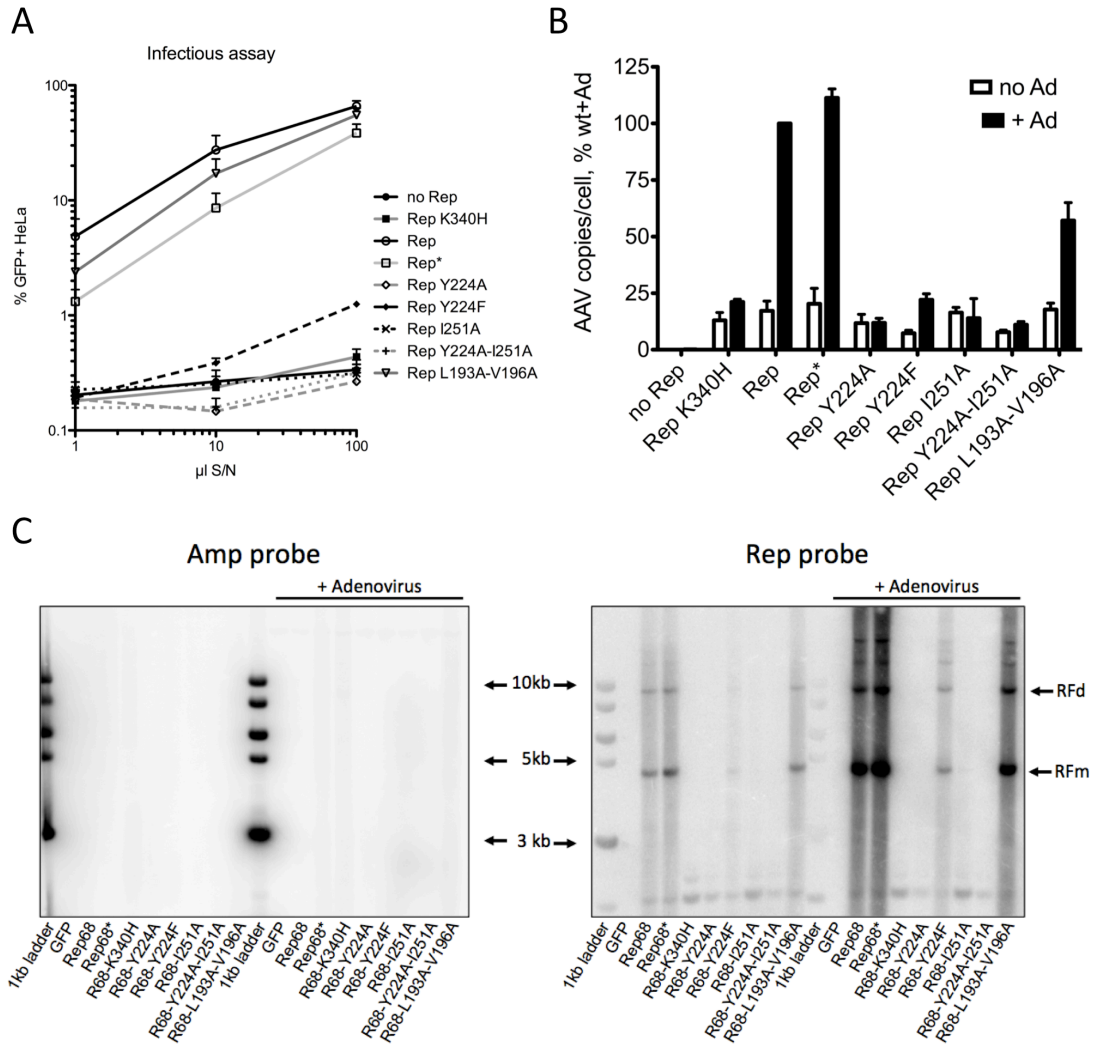


Figure 18: Oligomerisation-deficient mutants do not support the viral life cycle.
(A) The ability of the Rep68 interface mutants to support the AAV life cycle was assessed by producing rAAV2-GFP in presence of WT or mutant Rep68. Increasing volumes of supernatant collected from the producers cells was used to infect Hela cells. Data is presented as mean ± s.e.m. from three independent experiments. **(B)** AAV replication in presence of either Rep WT or the interface mutants was assessed by quantification of the amount of AAV DNA in 293T cells in presence of helper virus. **(C)** AAV replication was also assessed by analysing the formation of replicative intermediates in 293T cells in presence of helper virus. Left: Southern blot probed with a radiolabelled ampicillin probe to control for the digestion of input AAV DNA. Right, Southern blot probed with a radiolabelled Rep probe that detects replicated AAV DNA.

Taken together, these results suggest that oligomerisation of Rep proteins is essential for the efficient replication of AAV DNA and thus the production of infectious virus. This data, however, does not exclude the possibility that Rep

oligomerisation may be important for the packaging of viral DNA into preformed capsids.

5.2.4 The mutant Rep proteins are stable and localise to the nucleus

To exclude that the observed failure of the oligomerisation interface mutants to support AAV replication is due to a misfolded or unstable protein that is rapidly degraded, we analysed all mutants by western blot. In addition, we verified if the mutant proteins were able to translocate to the nucleus. Figure 19A shows that all mutants are expressed to similar levels 48h after transfection in 293T cells, suggesting that instability of the mutant proteins is not the cause of the observed lack of AAV replication. Similarly, all mutants show the expected nuclear localisation 24h post-transfection (Figure 19B).

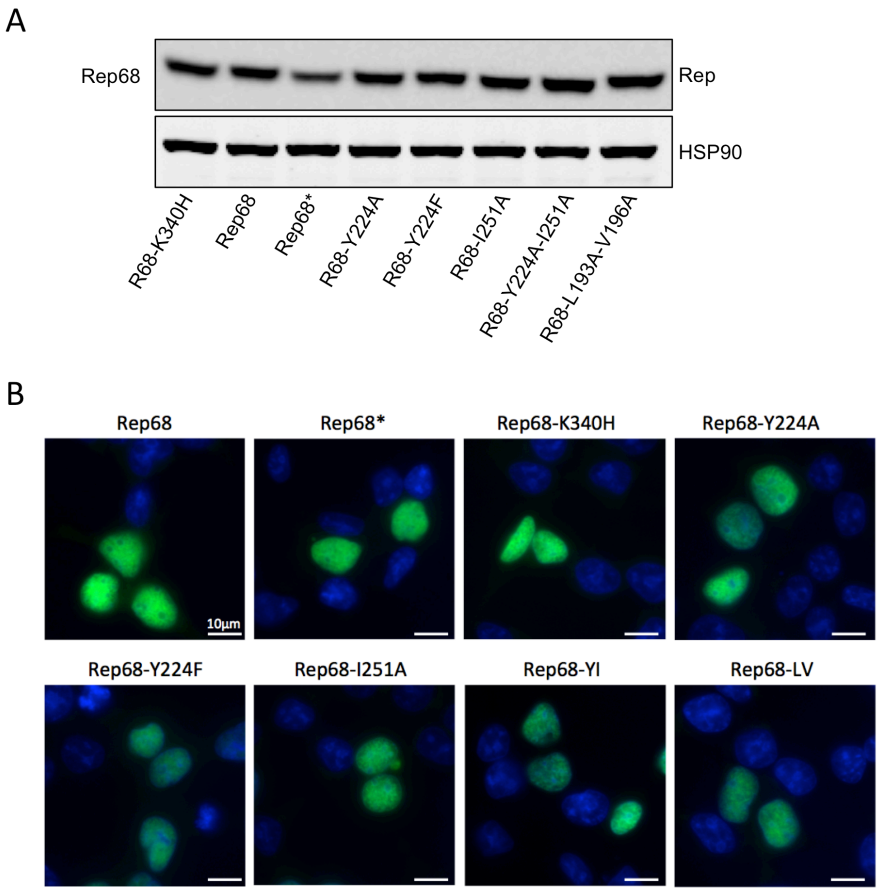


Figure 19: The interface mutants are stable and localise to the nucleus.
(A) Rep68 WT or mutant was transfected in 293T cells and was tested for protein expression levels by western blot 36h post-transfection. **(B)** 293T cells transfected with Rep68 WT or mutants were transfected and the Rep68 protein localisation was assessed 48h post-transfection by immunostaining (green). The nuclei are stained in blue.

5.2.5 Oligomerisation is required for DNA binding and nicking

In order to establish the cause of the observed phenotype of the oligomerisation-deficient mutants, we assessed their biochemical activities *in vitro*. Rep has three well-characterised enzymatic functions – RBS-specific DNA binding, TRS nicking and ATP-dependent DNA unwinding – all of which are necessary for AAV DNA replication. Table 1 shows the binding constants of Rep68* and mutant Rep proteins on p5 and *AAVS1* RBS-containing dsDNA substrates as determined by fluorescence anisotropy DNA binding assay. Rep68*, Rep68*-Y224F and the control Rep68*-K340H all appear to bind both DNA substrates efficiently. Rep68*-LV binds the RBS-containing DNA slightly less efficiently than WT, perhaps explaining the slightly lower levels of DNA replication observed (Figure 18B and C). We were not able to detect any binding for the other mutant Rep proteins, with the exception of Rep68*-I251A, which binds *AAVS1* DNA with 10-fold lower affinity than its WT counterpart.

Protein	AAVS1-41 [nM]	p5-41 [nM]
Rep68*	128	203
Rep68*-K340H	123	136
Rep68*-Y224A	nd*	nd*
Rep68*-Y224F	221	311
Rep68*-I251A	1438	nd*
Rep68*-YI	nd*	nd*
Rep68*-LV	281.3	795.5

Table 1: Oligomerisation-deficient mutants do not bind RBS-containing DNA substrates.
Two RBS-containing DNA substrates, one derived from *AAVS1* and the other from the p5 promoter, were used to calculate the DNA binding constants of Rep68* and interface mutants. Nd*: not determined. Courtesy of C. Escalante.

We then analysed the helicase activity of Rep68* and mutants on a fluorophore-labelled heteroduplex DNA substrate, which allows the monitoring of DNA unwinding. Surprisingly, all Rep68 mutants tested exhibited similar helicase activity, suggesting that Rep68 can unwind DNA under our experimental conditions even in

the absence of oligomerisation (Figure 20A). Alternatively, it is possible that a completely different oligomerisation interface is necessary for the helicase function.

Finally, we assessed the ability of the mutants to nick a supercoiled plasmid DNA containing RBS and TRS sequences (Figure 20B). A supercoiled plasmid containing the RBS and TRS region of *AAVS1* was previously shown to be efficiently nicked by Rep68 in an ATP-dependent manner (Lamartina et al. 2000a), in a reaction that closely mimics the terminal resolution step of AAV replication. Nicking by Rep68 at the TRS changes the plasmid conformation from supercoiled to open circular (Figure 20B), and the two plasmid species are easily distinguished by agarose gel electrophoresis. Because the nicking reaction follows the binding to the RBS site, it was no surprise that the mutants that failed to bind RBS-containing DNA also failed to nick the plasmid DNA (Figure 20C and D). Both the Rep68*-Y224F and the Rep68*-LV mutants were able to bind DNA, but while the Rep68*-LV mutant efficiently nicks the substrate Rep68*-Y224F showed significantly lower nicking activity compared to Rep68* (Figure 20C and D). This reduced nicking activity could explain the lower levels of AAV DNA replication observed for this mutant despite efficient RBS-containing DNA binding.

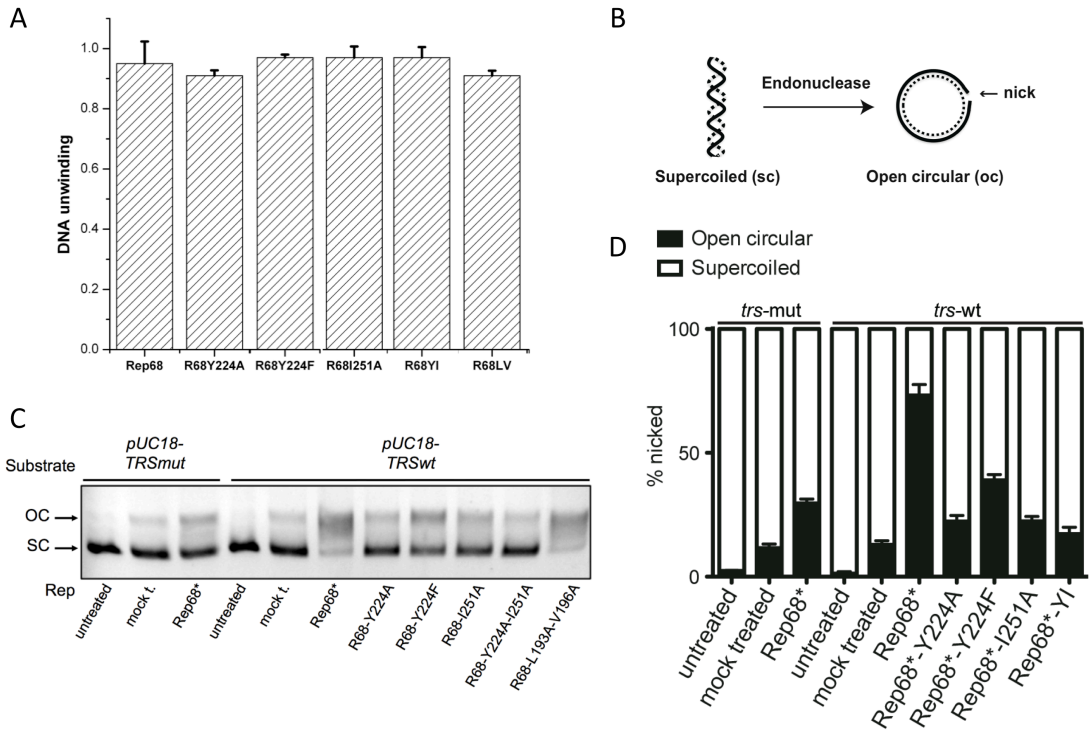


Figure 20: Biochemical characterisation of the interface mutants.
(A) The ability of the interface mutants to unwind DNA was assessed on a heteroduplex substrate. Courtesy of C. Escalante. **(B)** Schematic diagram of the mechanism behind the *in vitro* nicking assay. SC plasmid is mixed with the protein of interest; if the endonuclease activity is intact it will insert a nick and relax the plasmid conformation to an OC form. The SC and OC plasmid forms are readily separated by agarose gel electrophoresis. Figure courtesy of Leticia Agundez. **(C)** Representative example of the readout of this nicking assay. pUC18-TRSwt: pUC18-based plasmid containing the *AAVS1* sequences, including the RBS and TRS sites, that is readily nicked by Rep68. pUC18-TRSmut: this plasmid is identical to the nicking substrate pUC18-TRSwt except for a mutation in the TRS site that prevents nicking by Rep68. **(D)** Quantification of three independent experiments. Data presented as mean ± s.e.m.

5.2.6 Rep oligomerisation is important for transcriptional regulation

The dominant-negative mutant Rep-K340H, used here as a negative control for AAV replication, has been shown to fail to auto-regulate the levels of Rep proteins and to overproduce Rep proteins under permissive conditions (Chejanovsky and Carter 1990). To test whether oligomerisation also plays a role in the regulation of the levels of Rep proteins, we extracted the proteins from the cells used for the replication assay and assessed Rep proteins levels by immunoblotting. We observed a striking increase in the levels of all Rep isoforms in the presence of all the oligomerisation-defective mutant Rep proteins, as well as the control K340H mutant (Figure 21A, middle panel, Rep*-Y224A, Rep*-I251A, Rep*-YI). The Y224F

mutant, which partially oligomerises, showed a very modest increase in expression, while the LV double mutant showed Rep levels similar to the WT protein (Figure 21A, middle panel, Rep*-Y224F and Rep*-LV). Interestingly, Cap protein levels were the exact opposite: only in the presence of Rep proteins that support the AAV life cycle we detected high Cap levels, with the sole exception being the K340H mutant (Figure 21A, top panel). The same trend was observed both in the presence and absence of adenovirus infection, although the Cap levels were significantly lower in the absence of helper virus (Figure 21A, top panel). These results suggest that the oligomerisation mutants fail to regulate the levels of expression of the AAV proteins.

To better understand the important differences in protein amounts we assessed the levels of the AAV transcripts by RT-PCR (Figure 21B). Because all AAV mRNAs use the same polyadenylation signal, we could not quantify the p19 and p40 transcripts separately from the p5 transcripts using this technique. As expected, with all sets of primers used (one specific for p5 transcripts, one detects p5 and p19 transcripts and the last one detects all known AAV mRNAs), we observed an increase in mRNA levels in response to adenovirus in the presence of Rep proteins that support AAV replication (Figure 21B). In the presence of the Y224F mutant, the response to adenovirus is still present, but is nevertheless reduced compared to WT Rep. The oligomerisation-deficient mutants Y224A, I251A, and Y224A-I251A, which are also unable to bind the RBS at the p5 promoter (Table 1), have higher mRNA levels, varying between 2- and 10-fold, and do not respond to adenovirus infection. In the context of adenovirus co-infection, however, the differences in mRNA levels do not correlate with those of protein levels, suggesting that some level of post-transcriptional regulation is complicating the picture. In the presence of the K340H mutant, which oligomerises but fails to support AAV replication, we observed considerably higher basal mRNA levels compared to WT Rep and a lack of response to adenovirus, possibly explaining the high Cap protein amounts observed.

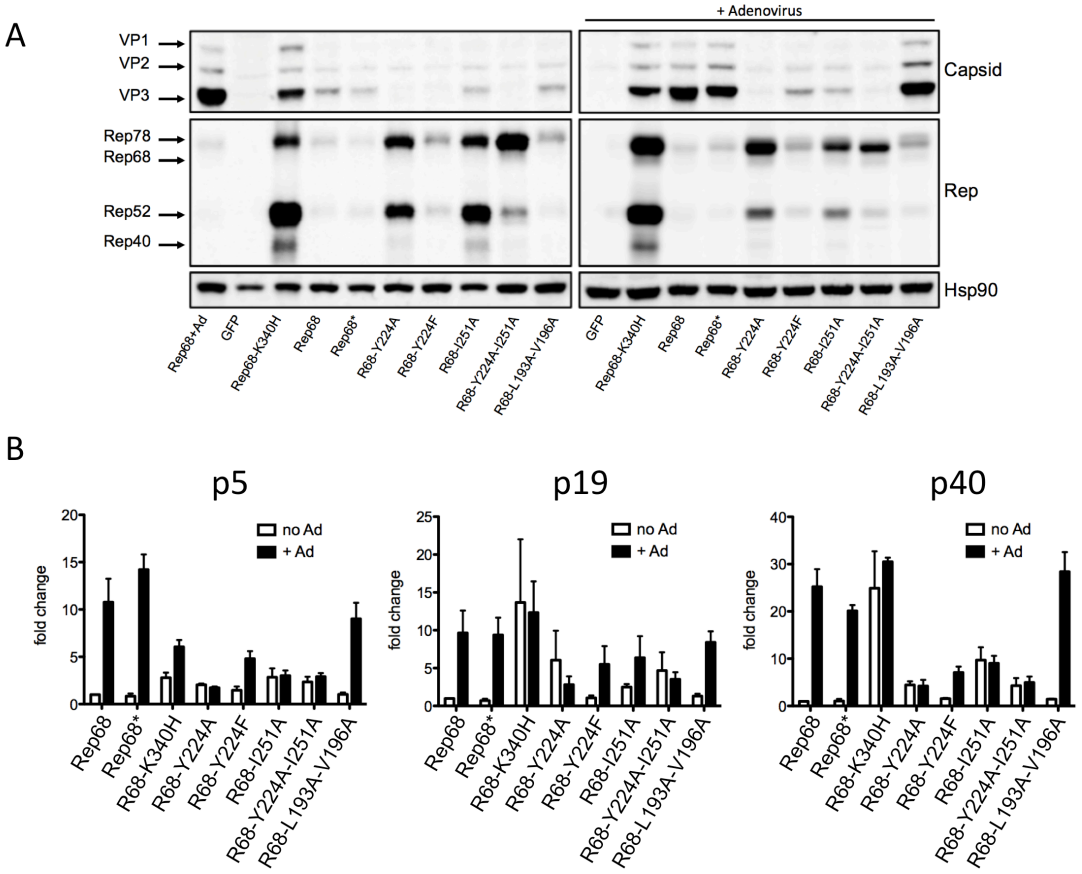


Figure 21: Rep oligomerisation is important for transcriptional regulation.
(A) Protein expression levels during permissive and non permissive conditions were assessed by transfecting an infectious plasmid containing the full AAV genome in 293T cells. The expression levels of the viral Rep and Cap proteins were detected by western blot. The first lane of the left panel is equivalent to the third lane of the right panel. **(B)** Transcription levels of AAV genes were investigated under the same experimental conditions. P5 (left panel) indicates that the data is specific for p5 transcripts; primers that bind after p19 but before 40 (middle panel) detect mRNAs generated from both p5 and p19 promoters; p40 primers detect mRNAs from all three viral promoters. The fold change was calculated relative to the mRNA levels in presence of WT Rep but in absence of adenovirus. Data is presented as mean ± s.e.m, N=3.

Taken together, our data supports a model in which Rep oligomerisation is involved in the regulation of the AAV promoters, potentially through binding at the RBS in the p5 promoter, and is necessary to achieve a correct transcription profile.

5.3 Discussion

The limited genome capacity of small viruses has driven the evolution of highly multifunctional non-structural proteins that combine several enzymatic functions necessary to support the viral life cycle. In AAV, the Rep proteins are responsible for orchestrating the whole viral life cycle, including transcriptional regulation,

replication, packaging of the viral DNA and Rep-mediated integration. The combination of several enzymatic functions, including DNA binding, nicking and unwinding, and the ability to interact with a multitude of DNA substrates and proteins, allows the Rep proteins to support the productive replication of AAV. The coordination of all these functions, thus, needs to be tightly controlled. We hypothesised that the substrate-dependent dynamic oligomerisation properties of AAV Rep contribute to the regulation of Rep functions. While it is clear that Rep oligomerisation is functionally relevant, little data is available on the regions of Rep that contribute to oligomerisation and data on possible oligomerisation interfaces are scarce. Initial studies that investigated the oligomerisation determinants identified the C-terminal region of Rep as important for multimerisation (Weitzman et al. 1996; Hermonat and Batchu 1997; Smith et al. 1997). Smith and colleagues used co-immunoprecipitation experiments to identify two regions that, when deleted, disrupted Rep oligomerisation: residues 151-188 forming a putative coiled-coil interaction, and residues 334-347 that include the ATP binding site (Smith et al. 1997). Importantly, in the hexamers of PV-E1 and SV40-LTag the corresponding ATP binding pocket is part of the oligomeric interface. The authors of this study went further and investigated how the point mutations L156G, L180T and K340H affected Rep oligomerisation. L156G showed a decrease but not a complete abrogation of Rep78-Rep78 binding, while the other mutants oligomerised normally (Smith et al. 1997). However, Smith and colleagues did not investigate the functional consequences of the oligomerisation-disrupting deletions and point mutations on Rep68. Davis et al., in a study that investigated the effects of charge cluster mutations on Rep68 functions, described several mutants that showed reduced oligomerisation as well as reduced functional activities (Davis et al. 1999). The authors placed particular emphasis on a putative coiled-coil domain located at the C-terminus of Rep68 and spanning residues 441 to 481, which included most of the mutations that affected oligomerisation. More recently, one report showed that residue R107, which is essential for RBS binding and nicking as well as for Rep-mediated plasmid integration (Urabe et al. 1999; Hickman et al. 2004), completely eliminates Rep68 oligomerisation in the presence of ssDNA (Mansilla-Soto et al. 2009). In this chapter I have described how, building on the most recent advances

in AAV2 Rep structural biology, we have designed new Rep mutants that are part of oligomeric interfaces in (potentially) functionally relevant Rep complexes. Contrary to previous studies, we have designed our mutants in regions that are not part of the well-characterised enzymatic folds of the Rep proteins, but rather fall in the recently described OD of Rep (Zarate-Perez et al. 2012).

The first oligomerisation interface was identified in a dimer model created using the available Rep40 structure with an extended N-terminal α-helix (up to residue 217) to complete the OD. In this model, a hydrophobic interface is formed through interactions of the OD of two Rep molecules that involve residues Y224 on one Rep molecule and I251 on the other (Figure 16A). I have described in the previous chapter that Y224 is important for Rep68 oligomerisation and its mutation to alanine disrupts the AAV life cycle. At that stage, however, the consequences of the Y224A mutation on specific Rep functions were not addressed. A similar but more conservative mutation, Y224F, was studied in an earlier report and was shown to completely disrupt DNA binding, endonuclease and helicase activities of a Rep68-MBP (maltose-binding protein) fusion *in vitro* (Walker et al. 1997a). To our knowledge, point mutations of residue I251 have not been studied before. The results presented here show that mutations Y224A, I251A and the combination of the two (Y224A-I251A) disrupt the oligomeric interface and that this interface is functionally relevant. While Rep68*-Y224A can still form some small multimers, Rep68*-I251 and Rep68*-YI are completely monomeric under our experimental conditions (Figure 17). These three oligomerisation-disrupting mutations completely abrogated the ability of Rep to support the production of infectious AAV-GFP particles (Figure 18A). Our data suggest that defects in DNA binding, and thus also in the endonuclease activity of Rep (Figure 20), are responsible for the absence of AAV DNA replication in the presence of these mutants and thus also for the lack of AAV-GFP particles production (Figure 18). Our data does not rule out the possibility that these mutants are also defective in viral DNA packaging. In addition, these mutants are also severely impaired in the regulation of the expression of the viral genes, and show increased protein levels of all Rep isoforms and lower levels of capsid proteins (Figure 21). The disruption of the gene-regulatory functions of

the Rep proteins, and their deregulated expression, could also contribute to the functional phenotypes observed.

The more conservative Rep-Y224F mutant had an intriguing intermediate phenotype. Analytical ultracentrifugation experiments suggest that Rep68*-Y224F retain the potential to form some of the oligomeric complexes formed by Rep68* *in vitro* (Figure 18). Our results indicate that this is not sufficient, however, to fully support the viral life cycle. The production of AAV-GFP particles and AAV replication in the presence of this mutant were severely affected, but not completely abolished, as was seen for the oligomerisation-deficient mutants (Figure 18). Surprisingly, and in disagreement with a previous report (Walker et al. 1997a), the binding of RBS-containing DNA does not appear to be compromised for this mutant (Figure 20A). A possible explanation for the observed defect in AAV replication is that the nicking activity of Rep68*-Y224F is severely impaired, although not entirely abolished (Figure 20B and C). Thus, this mutant retains the ability to bind RBS-containing DNA and to unwind DNA, but fails to promote the subsequent DNA nicking step. How this transition is affected, however, is not clear. An intriguing possibility would be that the initial Rep binding to the RBS and the origin melting promotes the assembly of a second Rep-DNA complex that is necessary for the nicking reaction, and that the Y224 residue plays a role in regulating this process. In contrast to our findings, the report by Walker and colleagues describes the Y224F mutants as completely defective in ITR binding, TRS endonuclease, DNA helicase and ATPase activities (Walker et al. 1997a). The differences between these studies are probably due to the different experimental strategies.

Surprisingly, all the mutants described here were still able to unwind a heteroduplex DNA substrate (Figure 20D). This suggests that interaction and unwinding of this DNA substrate do not require the formation of large Rep oligomers, or alternatively that the presence of heteroduplex DNA and ATP stabilises the formation of an oligomeric complex. Because Rep40 and Rep52 are monomeric in solution and only transiently form smaller oligomers in the presence of DNA and ATP but retain helicase activity (Dignam et al. 2007a; Zarate-Perez et al. 2012), it is plausible that the mutants described here are still able to unwind DNA by a mechanism equivalent to that of Rep52/40.

The regulation of protein levels appeared to be impaired in the presence of Rep mutants that do not bind p5 or AAVS1 DNA. The levels of Rep proteins are tightly regulated to be optimal for efficient AAV replication, and our results suggest that oligomerisation of the large Rep proteins is necessary for the correct regulation of the transcription of all AAV promoters (Figure 21). The oligomerisation-deficient mutants fail to induce transcription of the viral promoters upon infection with the helper virus adenovirus, and are also impaired in controlling the levels of both p5 and p19 transcripts. While the presence of the Y224A mutation creates a stronger Kozak sequence at the p19 promoter, this is not sufficient to explain the differences observed, in particular for the observed levels of Rep78. The increase in Rep52 expression in the presence of the Y224A-I251A mutations, in addition, is not as striking as that observed for the Y224A alone. We hypothesise that the defects observed are due to the lack of RBS binding by the oligomerisation-deficient mutants, rather than differences in the p19 promoter sequences.

The second interface we have characterised was directly observed in crystallography studies of the Rep OBD from AAV2 (Musayev et al. 2015a). In the AAV2 Rep*-OBD crystals, two OBD molecules are closely packed and interact through their C-terminal α-helix (α-helix F), forming a pseudo coiled-coil interaction (Figure 16C). The residues involved in the formation of this interface include residues L193 and V196 making hydrophobic interactions and residues H192 and S197 forming a hydrogen bond (Figure 16D). Because this OBD-OBD arrangement was also observed in studies of the AAV5 OBD (Hickman et al. 2004), we decided to investigate whether this interface was functionally relevant. The double mutant Rep*-L193A-V196A was found to be functional in all the assays used, and thus suggests that this pseudo coiled-coil interaction may be a crystallisation artefact that is not relevant for the AAV life cycle. Similarly, mutating residues H192 and S197 to alanine did not affect Rep oligomerisation and DNA binding (Musayev et al. 2015a).

In conclusion, our study identifies and describes two Rep-Rep protein interfaces. The results presented in this chapter suggest that one of these interfaces, stabilised by interactions that include the OD of Rep, is involved in the formation of Rep complexes and is functionally important throughout the AAV life cycle. Although we

have not tested the ability of interface mutants to mediate genomic integration, we anticipate that our oligomerisation-deficient mutants will not support this function. The identification of oligomeric interfaces, and further structural and functional characterisation of Rep oligomeric complexes, particularly in the presence of different DNA substrates, will provide additional insights into the molecular mechanisms behind the remarkable multifunctionality of AAV Rep proteins.

CHAPTER 6: REP-MEDIATED TARGETED GENE ADDITION IN CELLS WITH PROLIFERATIVE POTENTIAL

6.1 INTRODUCTION

To date, AAV remains the only known eukaryotic virus that has evolved a mechanism to site-specifically integrate in its host genome. This unique process is mediated by the large Rep proteins Rep78 and Rep68, which are versatile regulatory proteins that orchestrate the AAV life cycle. To improve our understanding of AAV Rep-mediated integration and its biotechnological potential, one approach is to characterise the mechanism of action of the Rep proteins on the *AAVS1* integration site and how the biochemical and structural properties of Rep78/68 mediate the integration event. In this context, the data presented in the first two chapters present an important platform that can be used for the characterisation of the complex formed by Rep78/68 on *AAVS1*. A second approach is to characterise the consequences of AAV integration on the host genome, including potential genotoxicity caused by the integration event. In this chapter, I present a project that was aimed at studying the consequences and the potential of Rep-mediated targeted gene addition in a clinically relevant disease model.

6.1.1 Gene therapy for monogenic immunodeficiency diseases

Primary immunodeficiencies (PIDs) are a heterogeneous group of genetic disorders that primarily affect the development and/or the functionality of the immune system (Fischer 2015). Common symptoms include an increased susceptibility to infections, autoimmunity and autoinflammatory diseases, as well as some malignancies. If left untreated, the most severe of these conditions, including X-linked severe combined immunodeficiency (SCID-X1), adenosine deaminase SCID (ADA-SCID), chronic granulomatous disease (CGD) and Wiskott-Aldrich syndrome (WAS), lead to early mortality in patients. The therapy of choice for severe PIDs is allogeneic transplantation of hematopoietic stem cells (HSCs) (Fischer 2015). The recent developments in transplantation protocols and in purification of HSCs have improved safety and efficacy of this treatment. However,

HLA-matched donors are still unavailable for a significant number of patients, making procedures based on *ex vivo* genetic correction of autologous HSCs clinically relevant (Rivat et al. 2012; Fischer et al. 2013). While AAV has been extensively developed as a gene therapy vector to treat non-dividing and differentiated cell populations, the episomal mode of rAAV persistence cannot be exploited for the treatment of PIDs, which require the genetic modification of stem cells and their proliferation to re-establish a functional immune system. More specifically, gene transfer to cells with proliferative potential such as HSCs requires transgene integration into the host genome to achieve lifelong expression. Thus, the vectors of choice for gene therapy treatment of PIDs have been based on retroviruses, which stably integrate their genome into the host genome. In particular, γ-retroviral vectors have been used in clinical trials for SCID-X1, ADA-SCID, CGD and WAS (Rivat et al. 2012; Fischer et al. 2013). Around 100 patients have now entered gene therapy clinical trials for the treatment of these PIDs, and the results have generally been encouraging, providing proof of principle that gene therapy can have long lasting curative effects and significantly improve the patient's life quality (Rivat et al. 2012; Fischer et al. 2013). Nonetheless, severe adverse events such as the development of T-cell leukaemia have been reported, highlighting the potential for insertional mutagenesis of retrovirus-mediated integration and clonal expansion (Williams and Thrasher 2014). In all cases, the development of leukaemia was associated with proviral integration events near proto-oncogenes such as the LMO2 gene and caused a deregulation of gene expression at the integration site, possibly due to the strong enhancer activity of the retroviral long terminal repeats (LTRs) (Williams and Thrasher 2014). More recent trials use newly designed retroviral vectors that have shown increased safety in pre-clinical and clinical studies, however the patient's haematopoietic clonal cell repertoires remain carefully monitored (Rivat et al. 2012; Fischer et al. 2013; Williams and Thrasher 2014).

6.1.2 AAV Rep-mediated targeted integration: an opportunity for targeted gene addition to the human genome?

Because of the intrinsically semi-random integration of retroviruses into the human genome, it is very difficult to predict the safety and efficacy of gene

therapies that use this approach to genetically manipulate cells with proliferative potential (Sadelain et al. 2012). In addition to the mentioned risk of insertional mutagenesis, the levels of expression of the transgene and its potential epigenetic silencing will vary depending on the chromosomal context surrounding the integration site (Sadelain et al. 2012). To overcome these issues, efforts have been made in recent years to develop strategies for the targeted modification of cells with proliferative potential, such as embryonic stem cells (ESCs), induced pluripotent stem cells (iPSCs) and HSCs. Examples include the use of bacteriophage integrases, homing nucleases, zinc-finger nucleases (ZFNs), transcription activator-like effector nucleases (TALENs) and most recently RNA-guided engineered nucleases (RGENs), based on the CRISPR-Cas system (Kim and Kim 2014). These strategies rely on the ability of the cellular machinery to repair double strand breaks (DSBs) in the DNA through homologous recombination. The nucleases can be engineered to introduce a DSB at a specific target in the genome, while exogenous donor DNA containing regions of homology to the targeted genomic locus provides the template for the repair of the double strand break by homologous recombination (Kim and Kim 2014). This mechanism can be exploited in several ways, including for repairing a mutated gene directly or for targeted transgene addition at a defined genomic locus. The efficiency and the specificity of this approach for targeted gene addition depend on the target locus, on the template DNA, and on the designer nuclease of choice (Kim and Kim 2014). In particular, genotoxicity due to off-target DSB induction and imperfect DSB repair is a significant risk that needs to be considered when using any of these approaches (Gabriel et al. 2011).

The preference of AAV for integration into *AAVS1* has prompted the investigation of AAV Rep-mediated integration as an alternative strategy for targeted gene addition. However, because *AAVS1* is located within a gene-dense region, assessing the genotoxicity of AAV integration is critical. Although this site has been designated as safe harbour site for transgene integration, the functional consequences of *AAVS1* disruption was initially only assessed in the context of AAV-mediated targeted integration (Henckaerts et al. 2009). Rep-mediated targeted insertion of a recombinant AAV vector encoding a marker gene into the *AAVS1*

ortholog in mouse embryonic stem cells (mESCs) resulted in strong transgene expression throughout *in vitro* differentiation into multiple lineages, without causing any apparent defects in differentiation (Dutheil et al. 2009; Henckaerts et al. 2009). Furthermore, these cells were able to successfully contribute to the development of all tissues when injected into blastocysts, without visible phenotypic defects throughout the life of the chimeric mice (Henckaerts et al. 2009). Similarly, Rep-mediated integration into *AAVS1* in human ESCs showed strong and stable marker gene expression, and did not affect the *in vitro* differentiation of the cells (Yang et al. 2008). In addition, these cells functionally integrated into a mouse heart when differentiated into human cardiac progenitor cells (Yang et al. 2008). Together, these experiments showed that Rep-mediated integration into *AAVS1* could result in safe and robust transgene addition in a translationally relevant context, and warranted further investigations of the feasibility of targeted gene addition by Rep-mediated integration.

Intriguingly, the AAV2 integration hotspot *AAVS1* has become one of the most used loci for targeted gene addition using designer nucleases and has proven to support robust transgene expression, including throughout ESC and iPSC differentiation (DeKelver et al. 2010; Lombardo et al. 2011; Sadelain et al. 2012; Tiyaboonchai et al. 2014).

6.1.3 Project outline and strategy

The aim of this project was to study the feasibility and safety profile of Rep-mediated targeted gene addition into *AAVS1* in a clinically relevant context. For this purpose, we decided to use SCID-X1 as a disease model. As described earlier in this chapter, proof of principle of the curative effects of gene therapy in SCID-X1 patients has been demonstrated using lentiviral vectors. SCID-X1 is the most common variant of SCID and results from loss of function mutations of the common cytokine receptor γ-chain, a shared component of the receptors for interleukin (IL)-2, IL-4, IL-7, IL-9, IL-15, and IL-21, encoded by the *IL2RG* gene (Fischer 2000). Because the *IL2RG* gene is located on the X-chromosome, the inheritance of this disease is sex-linked. Patients display a lack of T and NK cells, as well as abnormal B cell development; the disease is associated with early mortality if left untreated and

survival depends on the reconstitution of immunity by allogeneic transplantation or gene therapy (Fischer 2000). In the latter case, HSCs are isolated from the patient and transduced *ex vivo* with a retroviral vector carrying the *IL2RG* gene before being returned to the patient (Rivat et al. 2012; Fischer et al. 2013). AAV vectors, however, do not readily transduce HSCs. To overcome this problem, we took advantage of induced pluripotent stem cell technology to generate patient-specific iPSCs. Through our collaboration with Prof Adrian Thrasher of University College London, we were able to obtain SCID-X1 patient-derived dermal fibroblasts, which we reprogrammed to iPSCs to generate an unlimited source of SCID-X1 cells that could be transduced with AAV vectors and used for further studies. IPSCs are ESC-like pluripotent stem cells that are derived from somatic cells through forced expression of transcription factors such as Oct4, Klf4, Sox2 and c-Myc (Yamanaka 2007; Buganim et al. 2013). They were initially described as reprogramming factors by Takahashi and Yamanaka in 2006 (Takahashi and Yamanaka 2006), for which they received the Nobel Prize in 2012. In recent years the iPSC technology has greatly improved; iPSCs have been derived from a multitude of cell types, and several methods to deliver the reprogramming factors are now available (Gonzalez et al. 2011). These developments open the door to the production of autologous cells with proliferative potential for both cell therapy and drug screening (Inoue et al. 2014). In theory, it is now possible to obtain patient-specific cell lines for any disease and to subsequently use them to generate differentiated cell populations and to recapitulate the disease phenotype in vitro, allowing to screen for pathways that are involved in disease progression and/or therapies that can reverse the disease phenotype (Inoue et al. 2014).

As summarised in Figure 22, the first steps in this project were to obtain and characterise SCID-X1 iPSCs from the patient fibroblasts, and to design a rAAV-IL2RG vector to be used for targeted Rep-mediated gene addition into *AAVS1*. This would have been followed by a careful characterisation of the integration site(s) and surrounding regions, and, finally, by differentiation of the iPSCs into the hematopoietic lineage to compare the developmental potential of the isogenic SCID-X1 and corrected iPSCs. Unfortunately, this ambitious project was never

completed for the reasons explained below. In this chapter, I will present the initial developments of the project.

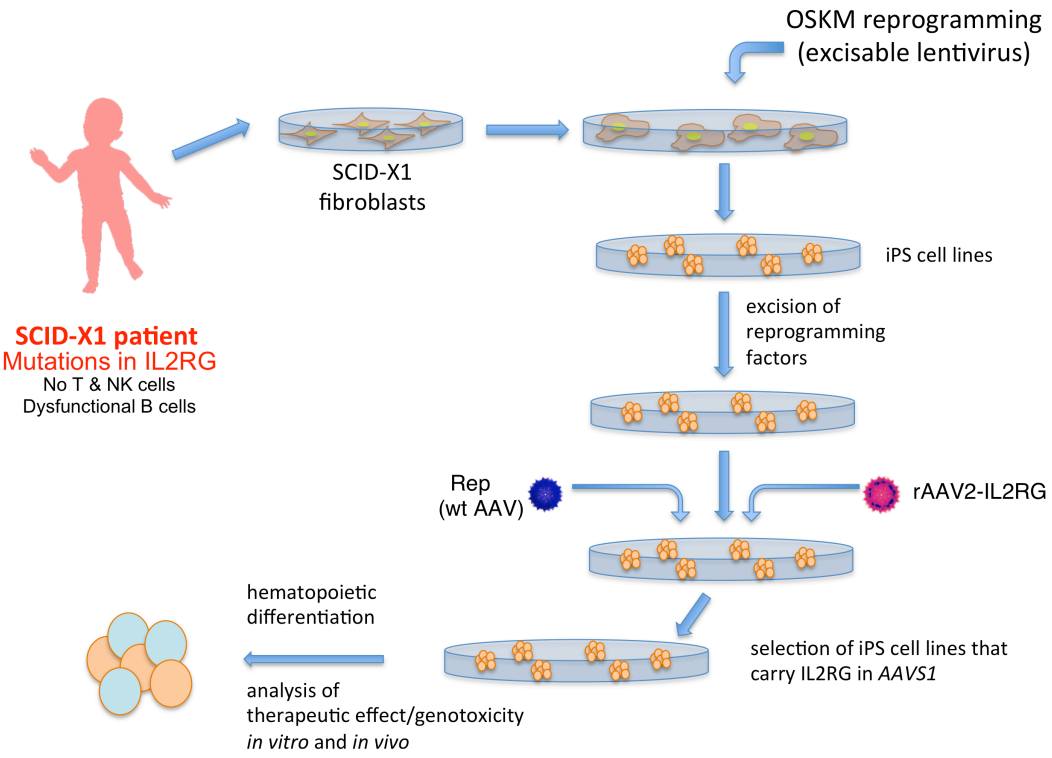


Figure 22: Project outline.
The first step of this project was the generation of patient-specific SCID-X1 iPSCs. The reprogramming factors would have been excised before infecting the iPSCs with a rAAV vector expressing the *IL2RG* gene and WT AAV2 to promote Rep-mediated integration in *AAVS1*.iPSC clones containing *IL2RG* integrated in *AAVS1* would then be selected and used to assess the consequences of integration throughout hematopoietic differentiation.

6.2 RESULTS

6.2.1 Vector design and plasmid validation

To produce our rAAV vector it was first necessary to clone the common γ-chain gene *IL2RG* into an ITR-containing plasmid. The initial vector design was based on the pTRUF11 construct (Zolotukhin et al. 1996). This plasmid contains the GFP gene under the control of the CAG promoter (a hybrid CMVie enhancer/chicken β-actin promoter) and neomycin resistance gene driven by the HSV thymidine kinase promoter, flanked by the AAV2 ITRs. The neomycin resistance cassette was replaced by a common γ-chain expression cassette containing a short version of the

EF-1 alpha promoter and the *IL2RG* cDNA, which was cloned from a lentiviral plasmid (kindly provided by Adrian Thrasher, Institute of Child Health, London, United Kingdom). Additionally, we designed a second construct that did not contain the GFP expression cassette. Before producing the AAV vectors, we assessed the γ-chain expression levels in transfected 293T cells by western blot and FACS (CD132) (Figure 23B and C, pTRUF11-GFP-IL2RG and pTRUF11-IL2RG). Compared to the original lentiviral plasmid (Figure 23B and C, SINLV-IL2RG) the γ-chain expression levels were very low, and thus we decided to design new *IL2RG* constructs. The new constructs were generated with the pSub201 plasmid, to take advantage of more suitable cloning sites and the reduced frequency of recombination of the ITRs (Samulski et al. 1987). We obtained and tested three new vectors. The first vector, pMB13, contained the same γ-chain expression cassette as described above but in the pSub201 backbone. To create the second construct, pSub201-*IL2RG*, we added the Woodchuck Hepatitis Virus Posttranscriptional Regulatory Element (WPRE) (Lee et al. 2005) behind the *IL2RG* gene to increase its expression (Figure 23A). The third plasmid was generated by adding a CMV-driven GFP expression cassette to pSub201-IL2RG (Figure 23A). Of the three new plasmids produced, pMB13 still showed low levels of γ-chain expression, while the two constructs that contained a WPRE element into the design (Figure 23A) expressed the γ-chain to levels comparable to those observed with the SINLV-IL2RG vector (Figure 23B and Table 2).

	% GFP	% CD132	% dpos
-ve ctrl	0.97	0.85	0.11
pTRUF11 (GFP ctrl)	65.1	0.73	14.1
pSINLV-IL2RG (CG ctrl)	0	82.2	0.42
pSUB201-IL2RG	0.07	68	0.38
pSUB201-IL2RG-GFP	11.9	7.45	71.7

Table 2: CD132 surface expression and GFP expression in 293Ts transfected with IL2RG-expressing plasmids.
The expression levels of the cell surface marker CD132 (γ-chain) and of GFP were assessed 48h after 293T transfection by FACS analysis. Values represent the mean of three experiments.

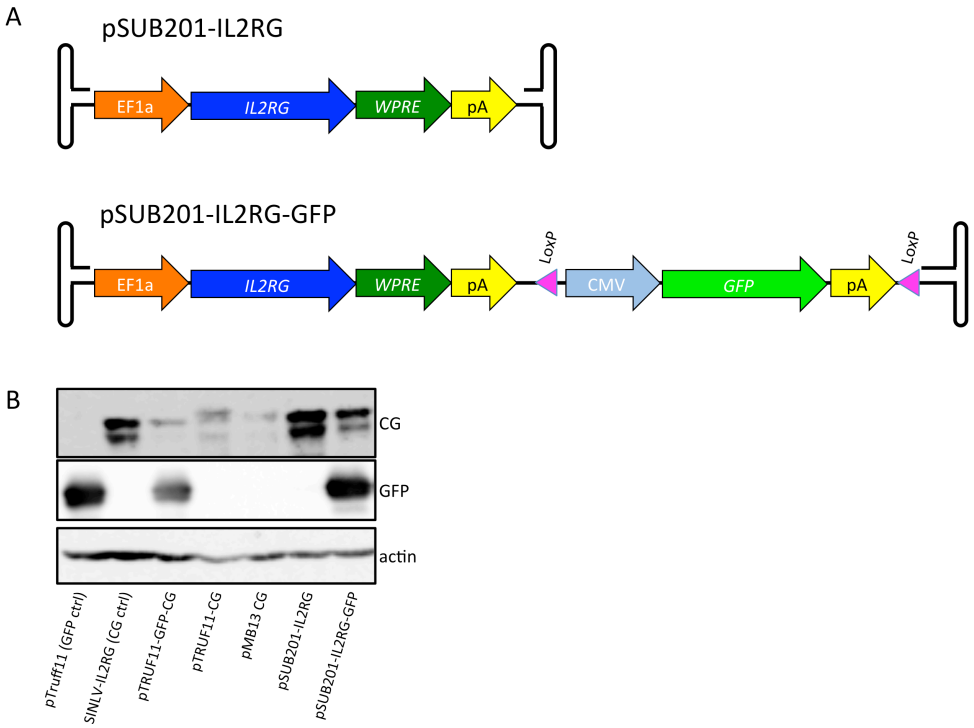


Figure 23: Vector design and validation of the plasmid constructs.
(A) Schematic design of the last two versions of IL2RG-expressing vectors used in this study. pSUB201-IL2RG contains the *IL2RG* expression cassette, controlled by the EF1α and WPRE regulatory elements. A second vector containing an excisable GFP expression cassette, pSUB201-IL2RG-GFP was designed to allow a simpler selection of transduced cells. **(B)** Western blot analysis of γ-chain and GFP expression levels from all the vector constructs built during this study 48h after transfection of 293T cells.

6.2.2 Vector production and viral vector testing

Based on our preliminary tests on the transgene-containing plasmids (Figure 23), we generated AAV2 vectors using the pSub201-IL2RG and the pSub201-IL2RG-GFP plasmids. The resulting AAV2-IL2RG and AAV2-IL2RG-GFP vectors were carefully titred: the viral particle titre was determined by quantifying the amount VP3 in each viral preparation by SDS-PAGE (Kohlbrenner et al. 2012) and the viral genome titre was measured by absolute qPCR quantification of *IL2RG* (Zeltner et al. 2010). The vectors were initially tested by transduction of 293T cells. Figure 24A shows the levels of γ-chain expression as determined by western blot and by CD132 surface staining (Table 3). While both vectors expressed the γ-chain, transduction with the vector co-expressing GFP resulted in significantly lower γ-chain expression levels. The GFP expression level was very high, suggesting some level of promoter

interference between the two expressions cassettes in our bicistronic vector (de Felipe 2002; Eszterhas et al. 2002). We used a cell line (ED7R-CG) that carries three integrated copies of the *IL2RG* gene (ED7R-CG, kindly provided by Prof Adrian Thrasher) as control for γ-chain expression.

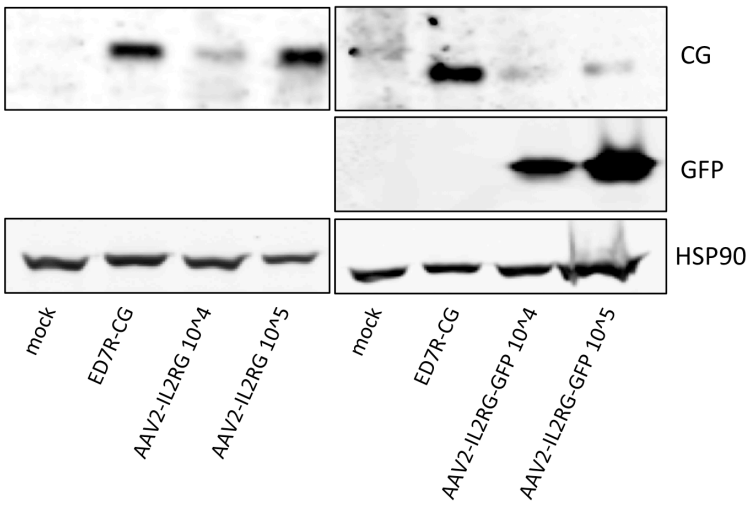


Figure 24: Western blot analysis of γ-chain and GFP expression levels.
293T cells were transduced with AAV2-IL2RG and AAV2-IL2RG-GFP AAV vectors and protein expression was assessed 48h after transduction.

condition	% GFP	% CD132	% dpos
-ve ctrl	0.9	0.8	0.07
ED7R-CG	0.7	44	2.3
AAV2-GFP	91.5	0.9	1.16
AAV2-IL2RG	0.07	10.7	0.38
AAV2-IL2RG-GFP	11.9	3.7	1.6

Table 3: CD132 surface expression and GFP expression in 293Ts transduced with *IL2RG*-expressing rAAVs.
The expression levels of the cell surface marker CD132 (γ-chain) and of GFP were assessed 48 hours after 293T transduction by FACS analysis. Values represent the mean of three experiments. ED7R-CG were used as positive control for the staining.

To further validate our vectors, we tested their ability to correct the ED7R cell line (kindly provided by Prof Adrian Thrasher), a human T-cell line that does not express the common γ-chain (Zhang et al. 2007). In addition to assessing the

expression levels of the CD132 surface marker, this cell line also allows for the study of the restoration of the IL2 receptor pathway by intracellular staining of phosphorylated STAT5, a downstream component of the IL2 signalling pathway. This pathway is functional in the ED7R-CG control cell line. Because AAV transduction of hematopoietic cells is generally poor, we first tested whether ED7R cells could be transduced with rAAV-GFP vectors, and found that AAV serotype 2 gave the highest transduction efficiency (data not shown). As shown in Table 4, transduction of ED7R cells with our *IL2RG* vectors did not result strong γ-chain expression and led to minimal functional activation of the IL2 signalling pathway. Transduction with AAV2-IL2RG-GFP and control AAV2-GFP vector, however, showed high GFP expression, indicating that transduction was not a limitation. Overall, these results suggest that expression of the IL2RG transgene in the context of the recombinant vector is sub-optimal in the ED7R cell line.

condition	% GFP	% CD132	% dpos	% P-STAT5
-ve ctrl	0.87	0.6	0.12	0.7
ED7R-CG	0.7	66	2.3	84.1
AAV2-GFP	90.9	0.71	0.42	1.38
AAV2-IL2RG	0.07	3.8	0.38	2.54
AAV2-IL2RG-GFP	80.9	1.39	1.5	2.29

Table 4: CD132 surface expression and IL2 pathway reconstitution in ED7R cells transduced with *IL2RG*-expressing rAAVs.
The expression levels of the cell surface marker CD132 (γ-chain), of GFP, and the level of ℗-STAT5 activation were assessed 48h after ED7R cells transduction by FACS analysis. Values represent the mean of two experiments. ED7R-CG were used as positive control for the staining.

6.2.3 Generation and characterisation of SCID-X1 iPSCs

While we were developing the AAV-IL2RG vectors, we started the generation of iPSCs from the SCID-X1 patient-derived dermal fibroblasts. For the generation of iPSCs, we used an excisable lentiviral vector containing the four reprogramming factors Oct4, Klf4, Sox2 and c-Myc (Somers et al. 2010) and a fibroblast reprogramming protocol kindly provided by Dr Paul Gadue (Children's Hospital of Philadelphia human ES/iPS cell core facility, Philadelphia, USA). Patient-derived dermal fibroblasts were transduced at passage 6 with varying MOIs of the

reprogramming retroviral vector. The day after transduction the culture medium was switched to a different formulation to improve the reprogramming efficiency. Five days post-transduction the cells were passaged and seeded on larger plates containing a feeder layer of irradiated mouse embryonic fibroblasts. The cells were left to grow without passaging for at least 3 weeks, after which the best-looking colonies were manually picked. Using this strategy we successfully generated and selected several iPSC clones and after 6 to 8 passages we proceeded to an initial assessment of pluripotency markers expression. Figure 25 shows a representative staining for the pluripotency markers nanog, Tra1-60 and SSEA4 of SCID-iPSCs clone 1. The SCID-X1 iPSCs showed pluripotency marker staining comparable to that of control hESCs. In addition, we tested the ability of different AAV serotypes to transduce the SCID-iPSCs using GFP expressing vectors. Table 5 shows that AAV2-GFP was the most efficient vector for the transduction of SCID-iPSCs.

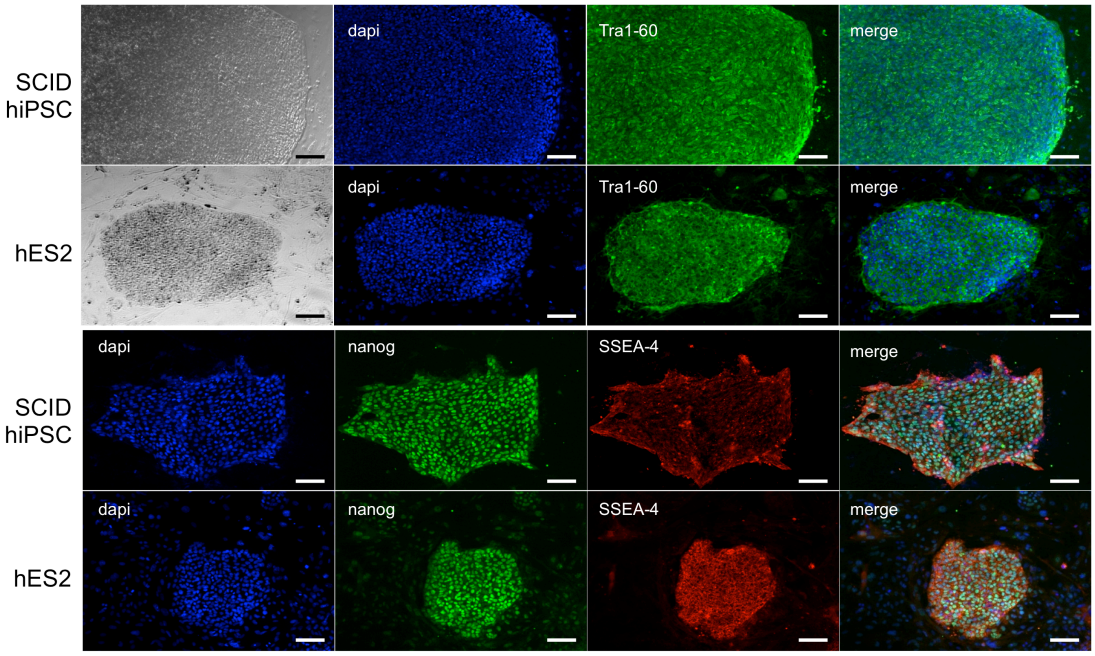


Figure 25: SCID-X1 iPSCs express pluripotency markers.
The expression of pluripotency markers Tra1-60, nanog, SSEA-4 was assessed by immunofluorescence. The hES2 ESC line was used as control. Scale bar is 100µl.

	AAV MOI 10^5 vg (%GFP)
neg ctrl	0.22
AAV1	5
AAV2	64.6
AAV5	0.5
AAV6	42.1
AAV8	14.4
AAV9	29.9

Table 5: Different AAV serotypes show different transduction efficiency of SCID-iPSCs.
The same GFP expression cassette was packaged in different AAV serotype capsids and was used to transduce SCID-iPSCs. Data from two experiments.

6.2.4 Problems and project termination

After the initial characterisation of the iPSCs obtained from SCID-X1 patient cells, the plan was to proceed with the excision of the reprogramming factors using the Cre-loxP system before transducing the cells with AAV. However, cultures that were initially looking healthy and displayed normal growth characteristics started to show changed morphology and drastically decreased cell viability. Cells tested negative for mycoplasma infection, a common cell culture contaminant that can have deleterious effects on stem cells (Young et al. 2010), and thus the reason for the sudden deterioration of the iPSCs cultures remained unknown. Control hESCs cultured in parallel in the same medium did not show signs of declining health. Finally, iPSCs that were left without media changes for 3 days tested positive for mycoplasma. After multiple attempts to establish the origin of the infection and rigorous testing of all the reagents, we identified the initial source of infection in the patient-derived fibroblasts, which presented a low-level mycoplasma contamination that remained undetectable with our luminescence-based mycoplasma testing method when the cell medium was changed daily. Mycoplasma infection, however, could be detected by using a more sensitive PCR-based detection technique or after prolonged culture without media changes (Young et al. 2010). Despite our efforts, all attempts to eradicate the infection either in the patient fibroblast or in the iPSCs using a curative regimen of anti-mycoplasma

antibiotics failed. Given that vector design would require further optimisation and that the process to receive ethical approval to obtain additional patient samples would take too long, we decided to abandon this project and focus our efforts on the Rep oligomerisation studies.

6.3 DISCUSSION

The aim of this project was to study the consequences and the potential of Rep-mediated targeted gene addition into *AAVS1* in a clinically relevant disease model such as SCID-X1. Our hope was to provide further insights into the development of an AAV-mediated technology for gene addition in cells with proliferative potential through site-specific integration. AAV is the only eukaryotic virus known to have evolved a mode of integration into the human genome that has a very strong preference for a single locus (*AAVS1*). As several aspects of the unique mechanism of integration of AAV are beginning to be understood (Henckaerts and Linden 2010), it is becoming increasingly tempting to suggest that this virus has evolved a strategy for minimal disruption of the host cell upon integration. Importantly, the observed duplication of the target gene, *PPP1R12C*, which is a consequence of the replication-based integration mechanism, could potentially allow for the 'removal' of the viral and duplicated cellular DNA, thereby restoring normal PPP1R12C expression. (Henckaerts et al. 2009).

Several questions, however, remain to be addressed. First of all, the frequency of Rep-mediated integration remains unclear. Studies that investigated the frequency of Rep-mediated integration describe different results, with frequencies that range from 0.1% to over 10% (Huser et al. 2002; Philpott et al. 2002; McCarty et al. 2004; Guilbaud et al. 2008; Henckaerts and Linden 2010). A comparison of these studies, however, is difficult due to the differences in the assays used to detect integration, the viral MOI, and the cell type used. In addition, several reports have recently challenged the actual specificity of Rep integration (Huser et al. 2010; Janovitz et al. 2013; Huser et al. 2014; Petri et al. 2015). In these studies, *AAVS1* was the most frequent hotspot for integration, but several other hotspots were found. Most likely due to the differences in experimental strategies, the identity of these additional integration sites was only partially overlapping amongst these

studies. AAV integration was shown to be enhanced by the presence of an RBS-like site and an appropriately spaced TRS homologue, suggesting that integration at non-*AAVS1* sites may also happen via a Rep-mediated mechanism (Huser et al. 2010; Janovitz et al. 2013; Huser et al. 2014; Petri et al. 2015). Achieving reasonably high frequency and high specificity of integration is a critical element to render a gene targeting technology safe and effective. Furthermore, the requirement for the presence of Rep78/68 for the mediation of integration into *AAVS1* represents an important technical challenge due to the pleiotropic effects of the Rep proteins. Safe and controlled Rep delivery strategies that induce a temporary and carefully balanced Rep expression to ensure efficient targeted integration combined with low cytotoxicity have yet to be fully developed (Recchia and Mavilio 2011). For the purpose of this project, we planned to use WT AAV alongside the rAAV-IL2RG vectors to provide Rep to mediate targeted integration, and to subsequently isolate cells that only contained integrated recombinant vector. If successful, our study could have served as proof-of-principle for the use of Rep-mediated targeted gene addition in cells with proliferative potential and would have warranted further efforts to develop this strategy for biotechnological purposes. For example, Rep could be provided *in trans* via vectors that do not have the ability to integrate, in which expression is controlled by a promoter that results in Rep expression levels similar to those observed in latent infection.

The most used strategies for genomic engineering include ZFNs, TALENs and RGENs, collectively referred to as programmable or designer nucleases (Kim and Kim 2014). The basic mechanism behind these three technologies is the same: the nuclease introduces a DSB into chromosomal DNA in a site-specific manner, thereby triggering endogenous DNA repair pathways, which ultimately result in genome modifications. While designer nucleases present several advantages and have been used efficiently in recent years, their reliance on DSB induction is not ideal and can result in significant genotoxicity (Kim and Kim 2014). While most DSBs will be repaired precisely by homologous recombination in the presence of homologous sequences, DSBs can also be repaired through the non-homologous end-joining (NHEJ) pathway, which will cause small deletions or insertions at the target site and lead to significant off target events (Gabriel et al. 2011). Rep, in

contrast, recognises a specific sequence in the human genome and induces a single-stranded nick at the cellular TRS site, avoiding the potentially deleterious effects caused by the induction of DSBs (Dutheil and Linden 2006). In addition, Rep-mediated integration has been suggested to lead to a partial duplication of the AAVS1 target site, potentially minimising the consequences of disruption of the *PPP1R12C* gene (Henckaerts et al. 2009). Interestingly, programmable nickases engineered from designer nucleases to only induce a ss nick instead of a DSB at the target site, have been shown to significantly lower genotoxicity and to promote a more precise genome editing, albeit at lower frequencies (Kim and Kim 2014). Ideally, a thorough study of the consequences of integration should take into account not only possible interferences with the expression of *PPP1R12C*, but also potential long-range effects on other genes, as well as consider the possibility of aberrant splicing events through cryptic splicing sites present in the vector genome as reported for lentiviral vector integrations (Cesana et al. 2012).

Zooming in on the results presented in this chapter, we were surprised to see a difference in transgene expression between our transfection and our transduction experiments. The reasons for this difference are not clear, but are likely to include differences in the number of gene copies available for transcription. This could possibly be caused by limited second-strand synthesis of our viral vectors, a known rate-limiting step for efficient rAAV transduction (Ferrari et al. 1996). This could be assessed by analysing the mRNA levels of our transgenes in transfected and transduced cells. In addition, the observation that GFP expression was much more robust than γ-chain expression points at significant differences in the strength of the expression cassettes used in our study, and possibly to some level of promoter interference (Eszterhas et al. 2002). Taken together, our results warrant a careful validation of vector design in studies that use AAV as their viral vector of choice.

This study was aimed to serve as proof-of-principle for Rep-mediated targeted gene addition in cells with proliferative potential in a clinically relevant model. If successful, it would have justified further efforts to develop this strategy for biotechnological purposes. At this time, however, the real potential of this technology remains unclear, and while Rep-mediated integration remains an important aspect of AAV biology worth investigating, the rapid developments of

more efficient and versatile technologies question the relevance of further studies of Rep-mediated gene addition as a technology. In particular, the discovery of the mechanism of action of the CRISPR-Cas9 system (Jinek et al. 2012) and its biotechnological development as highly efficient and highly specific RGENs (Sternberg and Doudna 2015), make it more and more challenging to invest in genome-modifying technologies that are significantly less efficient.

CHAPTER 7: GENERAL DISCUSSION AND PERSPECTIVES

AAV is a small, non-enveloped, ssDNA virus that is intrinsically replication incompetent and requires co-infection by a helper virus to replicate. It was initially discovered 50 years ago as a contaminant of adenoviral stocks. Despite 50 years of research and widespread infection throughout the human population, AAV has not been conclusively associated with any human pathology. Despite its small genome size, AAV has a tightly regulated life cycle that reflects its reliance on helper viruses for productive replication. In the presence of a helper virus, AAV DNA is efficiently replicated and packaged to form new infectious AAV particles. In the absence of helper virus, however, AAV can establish latency to persist in the infected cell. Two mechanisms have been described for latency: episomal persistence and chromosomal integration. These peculiar features of the AAV life style, combined with the lack of disease association and a certain ease of manipulation, have fuelled the development of AAV vectors for gene therapy and, today, most of the research on AAV revolves around its potential as a gene therapy vector. After pre-clinical trials showed encouraging safety and efficacy results in several disease models, numerous AAV vectors have now entered clinical trials, and several more are being developed towards clinical application. In particular, the AAV capsid has been the focus of detailed studies, and as capsid variants with defined tissue tropisms and specificities are becoming available, we are now closer to an AAV capsid toolkit that could allow tailoring of the capsid to the desired clinical target. Much remains to be learned from the basic AAV biology, however, and important contributions to AAV gene therapy are likely to come from basic research.

The biology of AAV presents several intriguing aspects that have yet to be fully uncovered. For example, AAV has evolved a single gene, called Rep, which encodes for the non-structural proteins that orchestrate the entire AAV life cycle. The products of the Rep gene, the proteins Rep78, Rep68, Rep52 and Rep40, are essential for virtually all stages of the AAV life cycle, including replication, packaging, transcriptional regulation, and chromosomal integration. The large Rep proteins, Rep78 and Rep68, are remarkable examples of viral multifunctional

proteins, which concentrate several enzymatic activities within a single polypeptide. Rep78/68 have DNA binding, nicking, helicase and ATPase activities, which are necessary to mediate several DNA transactions during the AAV life cycle, including the critical terminal resolution step of AAV replication and the integration of AAV into its chromosomal target site *AAVS1*. Rep78/68 have a multi-domain organisation: the N-terminal domain contains RBS-specific DNA binding and nicking activities, while the C-terminal domain is an ATP-dependent helicase. Despite the fact that the enzymatic functions of Rep78/68 have been well defined, less is known about the molecular mechanisms behind these functions. In particular, how the variety of enzymatic activities are regulated and coordinated remain elusive. The first objective of the research presented in this thesis was to contribute to the understanding of the structural requirements behind the biochemical activities of the large Rep proteins. More specifically, we aimed at defining functionally relevant determinants of Rep68 oligomerisation, which, we hypothesised, contribute to the coordination of Rep68 activities. Our investigations, presented in chapters 4 and 5, allowed us to identify a novel Rep68 mutant that prevented aggregation of this protein, which was hindering further structural studies of Rep68. This mutant has already proven its potential allowing a systematic study of Rep68 oligomerisation (Zarate-Perez et al. 2013), the first structural characterisation of the full-length Rep68 (Musayev et al. 2015a) and the crystallisation of the OBD of Rep68 in complex with the RBS at the *AAVS1* integration site (Musayev et al. 2015b). Furthermore, we defined a new oligomerisation domain (OD) in Rep68 that includes part of the interdomain linker and provides an elegant explanation of the differences in oligomerisation of the small and large Rep proteins. One linker residue, Y224, is conserved amongst SF3 helicases, and we showed that it plays a critical role in Rep oligomerisation and function. Finally, the results presented in chapter 5 describe the identification of an oligomeric interface that is essential for the correct functioning of Rep68. Disruption of this hydrophobic interface had drastic consequences on the biochemical activities of the Rep proteins and more generally on the AAV life cycle, supporting the hypothesis that Rep oligomerisation plays a critical role in the AAV life cycle. Determining how the large Rep proteins oligomerise could provide significant insights into the enzymatic mechanisms

behind Rep78/68 functions, including Rep-mediated chromosomal integration. Further studies on the postulated role of substrate-dependent oligomerisation in coordinating Rep78/68 enzymatic activities, in particular, will be important.

Identifying similarities and differences in the oligomeric properties of proteins related to Rep78/68, such as parvoviral NS1 proteins, SF3 helicases or HUH endonucleases, could also foster interesting developments. For instance, the functional conservation of structural motifs within the HUH family of endonucleases or within the SF3 helicases, could be explored by swapping catalytic motifs. This approach has been used previously to demonstrate that the *ori*-specific DNA binding activity of Rep78/68 was fully contained within its OBD (Cathomen et al. 2000; Yoon et al. 2001). Yoon and colleagues, however, were also able to demonstrate that a chimeric protein made by combining the OBD of the goose parvovirus (GPV) Rep1 protein and the helicase domain of AAV2 Rep78 fully supported the *in vitro* replication of a substrate that contained the GPV *ori*. It would be interesting to test whether this chimeric protein supports GPV replication in cells that are normally infected by GPV and in cells that are specific for AAV2 replication. It would be even more intriguing to assess the ability of less related HUH endonucleases, such as plant viruses or even bacterial relaxases, to functionally combine with the helicase domain of Rep. Potentially, this would allow to redirect protein functions from one species to another, and could expand the number of available substrates for enzymes that have biotechnological applications. Understanding the precise domain boundaries, as well as defining the oligomeric properties that support the functions investigated, is critical for the informed design of such studies. For example, it is likely that functionally combining the functions of two proteins that form the same oligomeric complex will be easier than combining domains from proteins that have distinct oligomeric behaviours. Importantly, the apparent conservation of an OD in the region that connects catalytically active domains, such as the OBD and helicase domains of Rep68, needs to be carefully considered in designing chimeric proteins whose oligomeric properties are assumed to be functionally relevant. In the context of Rep68 functions in the AAV life cycle, it would be intriguing to investigate the possibility to

redirect the specificity of Rep-mediated integration into the human genome to different target sites by combining it with the OBD of a related protein.

Rep-mediated integration into the integration site *AAVS1* was the focus of the last chapter of this thesis. Our aim was to study the safety and the feasibility of a Rep-mediated strategy to site-specifically add a transgene into the safe-harbour locus *AAVS1* in a translationally relevant disease model, namely SCID-X1 iPSCs. The standard cure for SCID-X1, a devastating immunodeficiency, is allogeneic bone marrow transplantation, but when a patient lacks an HLA-matched donor, gene therapy is a valuable option. Clinical trials that used autologous HSCs transduced with a retroviral vector harbouring the corrective gene to reconstitute immunity in SCID-X1 patients have shown great promise. A small percentage of the patients treated with this approach, however, developed significant side effects in the form of leukaemia due to the activation of proto-oncogenes by the integrated provirus. These results warranted the investigation of alternative strategies to modify the genome in a controlled way. It is in this context that we decided to assess if Rep mediated integration would be a viable alternative approach to modify cells with proliferative potential. Our hypothesis was that Rep-mediated integration into *AAVS1* is a naturally evolved strategy to site-specifically insert exogenous genetic material into the human genome without major consequences for the host cell. Our hypothesis was based on reports that showed that human and mouse ESCs carrying a transgene integrated into *AAVS1* by Rep-mediated integration showed normal differentiation *in vitro* and *in vivo* and robust transgene expression throughout differentiation. After successfully generating iPSCs from SCID-X1 fibroblasts and validating a potential therapeutic targeting rAAV vector, unfortunately, an untreatable mycoplasma infection of the iPSCs forced us to terminate this project. If successful, this project would have supported the development of Rep-mediated integration as a viable approach to achieve targeted gene addition in clinically relevant cell populations. However, during the time of this thesis, important technological advancements, such as the development of RGENs as well as significant contributions to our understanding of AAV integration, and particularly of its specificity, argue against the development of Rep-mediated integration as a biotechnological tool. The low efficiency and the newly described limited specificity

of AAV integration are significant limitations that cannot be overlooked. Only developments that will overcome these important limitations would permit Rep-mediated integration to become an attractive option for gene therapy. The study of the importance of AAV integration into the human genome, however, is still relevant in the context of the biology of the AAV life cycle. Investigating the existence of integrated AAV sequences *in vivo* will be particularly relevant to the establishment of the role of Rep-mediated integration in the AAV life cycle.

Finally, in addition to the research questions highlighted here and discussed throughout the rest of this thesis, two facets of the basic AAV biology may see significant developments in the coming years. The first is the study of AAV interactions with the host cell and the consequences of these interactions on the AAV life cycle. Most of the information available to date is extremely fragmented, with multiple individual reports that identify a role for cellular factors but that have not been followed up. Defining a coherent picture of the role of pathways of innate immunity or of DDR, for example will provide intriguing insights into AAV biology and the reasons for AAV's inability to replicate autonomously. Furthermore, investigating potential restriction factors for AAV replication, gene expression or packaging, could also improve the current protocols for the production of AAV vectors. As these host-virus interactions become better defined and a role for the Rep proteins in these interactions is established, structure-function studies and biophysical analyses as those presented here will further this field. The second aspect of AAV biology that is particularly intriguing and warrant further investigations is the dependence of AAV on helper virus functions. It is becoming increasingly clear that viruses interact not only with their host but also with each other. Improving our understanding of AAV-adenovirus and AAV-HSV interactions may reveal important features of interviral interactions that could be applied to other viral systems, including clinically relevant ones.

BIBLIOGRAPHY

Abbate, E. A., J. M. Berger, et al. (2004). "The X-ray structure of the papillomavirus helicase in complex with its molecular matchmaker E2." Genes Dev **18**(16): 1981-1996.

Agbandje-McKenna, M. and J. Kleinschmidt (2011). "AAV capsid structure and cell interactions." Methods Mol Biol **807**: 47-92.

Alazard-Dany, N., A. Nicolas, et al. (2009). "Definition of herpes simplex virus type 1 helper activities for adeno-associated virus early replication events." PLoS Pathog **5**(3): e1000340.

Astell, C. R., C. D. Mol, et al. (1987). "Structural and functional homology of parvovirus and papovavirus polypeptides." J Gen Virol **68 (Pt 3)**: 885-893.

Atchison, R. W., B. C. Casto, et al. (1965). "Adenovirus-Associated Defective Virus Particles." Science **149**(3685): 754-756.

Balague, C., M. Kalla, et al. (1997). "Adeno-associated virus Rep78 protein and terminal repeats enhance integration of DNA sequences into the cellular genome." J Virol **71**(4): 3299-3306.

Bantel-Schaal, U., I. Braspenning-Wesch, et al. (2009). "Adeno-associated virus type 5 exploits two different entry pathways in human embryo fibroblasts." J Gen Virol **90**(Pt 2): 317-322.

Bantel-Schaal, U. and H. zur Hausen (1984). "Characterization of the DNA of a defective human parvovirus isolated from a genital site." Virology **134**(1): 52-63.

Bar, S., J. Rommelaere, et al. (2013). "Vesicular transport of progeny parvovirus particles through ER and Golgi regulates maturation and cytolysis." PLoS Pathog **9**(9): e1003605.

Batchu, R. B., R. M. Kotin, et al. (1994). "The regulatory rep protein of adeno-associated virus binds to sequences within the c-H-ras promoter." Cancer Lett **86**(1): 23-31.

Beaton, A., P. Palumbo, et al. (1989). "Expression from the adeno-associated virus p5 and p19 promoters is negatively regulated in trans by the rep protein." J Virol **63**(10): 4450-4454.

Bergvall, M., T. Melendy, et al. (2013). "The E1 proteins." Virology **445**(1-2): 35-56.

Berns, K. I. and T. J. Kelly, Jr. (1974). "Letter: Visualization of the inverted terminal repetition in adeno-associated virus DNA." J Mol Biol **82**(2): 267-271.

Berthet, C., K. Raj, et al. (2005). "How adeno-associated virus Rep78 protein arrests cells completely in S phase." Proc Natl Acad Sci U S A **102**(38): 13634-13639.

Blacklow, N. R., M. D. Hoggan, et al. (1970). "Adenovirus-associated viruses: enhancement by human herpesviruses." Proc Soc Exp Biol Med **134**(4): 952-954.

Blacklow, N. R., M. D. Hoggan, et al. (1967). "Isolation of adenovirus-associated viruses from man." Proc Natl Acad Sci U S A **58**(4): 1410-1415.

Blacklow, N. R., M. D. Hoggan, et al. (1968). "Serologic evidence for human infection with adenovirus-associated viruses." J Natl Cancer Inst **40**(2): 319-327.

Bleker, S., M. Pawlita, et al. (2006). "Impact of capsid conformation and Rep-capsid interactions on adeno-associated virus type 2 genome packaging." J Virol **80**(2): 810-820.

Boer, D. R., J. A. Ruiz-Maso, et al. (2009). "Plasmid replication initiator RepB forms a hexamer reminiscent of ring helicases and has mobile nuclease domains." EMBO J **28**(11): 1666-1678.

Bohenzky, R. A., R. B. LeFebvre, et al. (1988). "Sequence and symmetry requirements within the internal palindromic sequences of the adeno-associated virus terminal repeat." Virology **166**(2): 316-327.

Brister, J. R. and N. Muzyczka (1999). "Rep-mediated nicking of the adeno-associated virus origin requires two biochemical activities, DNA helicase activity and transesterification." J Virol **73**(11): 9325-9336.

Brister, J. R. and N. Muzyczka (2000). "Mechanism of Rep-mediated adeno-associated virus origin nicking." J Virol **74**(17): 7762-7771.

Buganim, Y., D. A. Faddah, et al. (2013). "Mechanisms and models of somatic cell reprogramming." Nat Rev Genet **14**(6): 427-439.

Calcedo, R., L. H. Vandenberghe, et al. (2009). "Worldwide epidemiology of neutralizing antibodies to adeno-associated viruses." J Infect Dis **199**(3): 381-390.

Carter, B. J. (2004). "Adeno-associated virus and the development of adeno-associated virus vectors: a historical perspective." Mol Ther **10**(6): 981-989.

Carter, B. J., G. Khoury, et al. (1975). "Physical map and strand polarity of specific fragments of adenovirus-associated virus DNA produced by endonuclease R-EcoRI." J Virol **16**(3): 559-568.

Carter, B. J., G. Khoury, et al. (1972). "Adenovirus-associated virus multiplication. IX. Extent of transcription of the viral genome in vivo." J Virol **10**(6): 1118-1125.

Cassell, G. D. and M. D. Weitzman (2004). "Characterization of a nuclear localization signal in the C-terminus of the adeno-associated virus Rep68/78 proteins." Virology **327**(2): 206-214.

Cathomen, T., D. Collete, et al. (2000). "A chimeric protein containing the N terminus of the adeno-associated virus Rep protein recognizes its target site in an in vivo assay." J Virol **74**(5): 2372-2382.

Cervelli, T., J. A. Palacios, et al. (2008). "Processing of recombinant AAV genomes occurs in specific nuclear structures that overlap with foci of DNA-damage-response proteins." J Cell Sci **121**(Pt 3): 349-357.

Cesana, D., J. Sgualdino, et al. (2012). "Whole transcriptome characterization of aberrant splicing events induced by lentiviral vector integrations." J Clin Invest **122**(5): 1667-1676.

Chandler, M., F. de la Cruz, et al. (2013). "Breaking and joining single-stranded DNA: the HUH endonuclease superfamily." Nat Rev Microbiol **11**(8): 525-538.

Chang, L. S., Y. Shi, et al. (1989). "Adeno-associated virus P5 promoter contains an adenovirus E1A-inducible element and a binding site for the major late transcription factor." J Virol **63**(8): 3479-3488.

Chejanovsky, N. and B. J. Carter (1989). "Mutagenesis of an AUG codon in the adeno-associated virus rep gene: effects on viral DNA replication." Virology **173**(1): 120-128.

Chejanovsky, N. and B. J. Carter (1990). "Mutation of a consensus purine nucleotide binding site in the adeno-associated virus rep gene generates a dominant negative phenotype for DNA replication." J Virol **64**(4): 1764-1770.

Chen, C. L., R. L. Jensen, et al. (2005). "Molecular characterization of adeno-associated viruses infecting children." J Virol **79**(23): 14781-14792.

Cheung, A. K., M. D. Hoggan, et al. (1980). "Integration of the adeno-associated virus genome into cellular DNA in latently infected human Detroit 6 cells." J Virol **33**(2): 739-748.

Chiorini, J. A., S. Afione, et al. (1999a). "Adeno-associated virus (AAV) type 5 Rep protein cleaves a unique terminal resolution site compared with other AAV serotypes." J Virol **73**(5): 4293-4298.

Chiorini, J. A., F. Kim, et al. (1999b). "Cloning and characterization of adeno-associated virus type 5." J Virol **73**(2): 1309-1319.

Chiorini, J. A., M. D. Weitzman, et al. (1994a). "Biologically active Rep proteins of adeno-associated virus type 2 produced as fusion proteins in Escherichia coli." J Virol **68**(2): 797-804.

Chiorini, J. A., S. M. Wiener, et al. (1994b). "Sequence requirements for stable binding and function of Rep68 on the adeno-associated virus type 2 inverted terminal repeats." J Virol **68**(11): 7448-7457.

Chiorini, J. A., L. Yang, et al. (1997). "Cloning of adeno-associated virus type 4 (AAV4) and generation of recombinant AAV4 particles." J Virol **71**(9): 6823-6833.

Chirmule, N., K. Propert, et al. (1999). "Immune responses to adenovirus and adeno-associated virus in humans." Gene Ther **6**(9): 1574-1583.

Choi, V. W., D. M. McCarty, et al. (2006). "Host cell DNA repair pathways in adeno-associated viral genome processing." J Virol **80**(21): 10346-10356.

Clement, N., D. R. Knop, et al. (2009). "Large-scale adeno-associated viral vector production using a herpesvirus-based system enables manufacturing for clinical studies." Hum Gene Ther **20**(8): 796-806.

Collaco, R. F., V. Kalman-Maltese, et al. (2003). "A biochemical characterization of the adeno-associated virus Rep40 helicase." J Biol Chem **278**(36): 34011-34017.

Costello, E., P. Saudan, et al. (1997). "High mobility group chromosomal protein 1 binds to the adeno-associated virus replication protein (Rep) and promotes Rep-mediated site-specific cleavage of DNA, ATPase activity and transcriptional repression." EMBO J **16**(19): 5943-5954.

Crawford, L. V., E. A. Follett, et al. (1969). "The DNA of a minute virus of mice." J Gen Virol **4**(1): 37-46.

D'Amours, D. and S. P. Jackson (2002). "The Mre11 complex: at the crossroads of dna repair and checkpoint signalling." Nat Rev Mol Cell Biol **3**(5): 317-327.

Davis, M. D., R. S. Wonderling, et al. (1999). "Analysis of the effects of charge cluster mutations in adeno-associated virus Rep68 protein in vitro." J Virol **73**(3): 2084-2093.

Davis, M. D., J. Wu, et al. (2000). "Mutational analysis of adeno-associated virus type 2 Rep68 protein endonuclease activity on partially single-stranded substrates." J Virol **74**(6): 2936-2942.

de Felipe, P. (2002). "Polycistronic viral vectors." Curr Gene Ther **2**(3): 355-378.

DeKelver, R. C., V. M. Choi, et al. (2010). "Functional genomics, proteomics, and regulatory DNA analysis in isogenic settings using zinc finger nuclease-driven transgenesis into a safe harbor locus in the human genome." Genome Res **20**(8): 1133-1142.

Denhardt, D. T., S. Eisenberg, et al. (1976). "Multiple structures of adeno-associated virus DNA: analysis of terminally labeled molecules with endonuclease R-Hae III." J Virol **18**(2): 672-684.

Di Pasquale, G. and S. N. Stacey (1998). "Adeno-associated virus Rep78 protein interacts with protein kinase A and its homolog PRKX and inhibits CREB-dependent transcriptional activation." J Virol **72**(10): 7916-7925.

Dignam, S. S., R. F. Collaco, et al. (2007a). "Coupled ATP and DNA binding of adeno-associated virus Rep40 helicase." Biochemistry **46**(2): 568-576.

Dignam, S. S., J. J. Correia, et al. (2007b). "Activation of the ATPase activity of adeno-associated virus Rep68 and Rep78." Biochemistry **46**(21): 6364-6374.

DiMattia, M. A., H. J. Nam, et al. (2012). "Structural insight into the unique properties of adeno-associated virus serotype 9." J Virol **86**(12): 6947-6958.

Drouin, L. M. and M. Agbandje-McKenna (2013). "Adeno-associated virus structural biology as a tool in vector development." Future Virol **8**(12): 1183-1199.

Durocher, Y., P. L. Pham, et al. (2007). "Scalable serum-free production of recombinant adeno-associated virus type 2 by transfection of 293 suspension cells." J Virol Methods **144**(1-2): 32-40.

Dutheil, N., E. Henckaerts, et al. (2009). "Transcriptional analysis of the adeno-associated virus integration site." J Virol **83**(23): 12512-12525.

Dutheil, N. and R. M. Linden (2006). Site-specific integration by adeno-associated virus. Parvoviruses. J. R. Kerr, S. F. Cotmore, M. E. Bloom, R. M. Linden and C. R. Parrish. London, Hodder Arnold**:** 214-236.

Dutheil, N., F. Shi, et al. (2000). "Adeno-associated virus site-specifically integrates into a muscle-specific DNA region." Proc Natl Acad Sci U S A **97**(9): 4862-4866.

Dutheil, N., S. C. Smith, et al. (2014). "Adeno-associated virus Rep represses the human integration site promoter by two pathways that are similar to those required for the regulation of the viral p5 promoter." J Virol **88**(15): 8227-8241.

Dutheil, N., M. Yoon-Robarts, et al. (2004). "Characterization of the mouse adeno-associated virus AAVS1 ortholog." J Virol **78**(16): 8917-8921.

Enemark, E. J. and L. Joshua-Tor (2006). "Mechanism of DNA translocation in a replicative hexameric helicase." Nature **442**(7100): 270-275.

Erles, K., P. Sebokova, et al. (1999). "Update on the prevalence of serum antibodies (IgG and IgM) to adeno-associated virus (AAV)." J Med Virol **59**(3): 406-411.

Eszterhas, S. K., E. E. Bouhassira, et al. (2002). "Transcriptional interference by independently regulated genes occurs in any relative arrangement of the genes and is influenced by chromosomal integration position." Mol Cell Biol **22**(2): 469-479.

Fanning, E. and K. Zhao (2009). "SV40 DNA replication: from the A gene to a nanomachine." Virology **384**(2): 352-359.

Ferrari, F. K., T. Samulski, et al. (1996). "Second-strand synthesis is a rate-limiting step for efficient transduction by recombinant adeno-associated virus vectors." J Virol **70**(5): 3227-3234.

Fischer, A. (2000). "Severe combined immunodeficiencies (SCID)." Clin Exp Immunol **122**(2): 143-149.

Fischer, A. (2015). "Recent advances in understanding the pathophysiology of primary T cell immunodeficiencies." Trends Mol Med **21**(7): 408-416.

Fischer, A., S. Hacein-Bey-Abina, et al. (2013). "Gene therapy of primary T cell immunodeficiencies." Gene **525**(2): 170-173.

Flotte, T. R., S. A. Afione, et al. (1993). "Stable in vivo expression of the cystic fibrosis transmembrane conductance regulator with an adeno-associated virus vector." Proc Natl Acad Sci U S A **90**(22): 10613-10617.

Flotte, T. R., X. Barraza-Ortiz, et al. (1995). "An improved system for packaging recombinant adeno-associated virus vectors capable of in vivo transduction." Gene Ther **2**(1): 29-37.

Flotte, T. R., P. L. Zeitlin, et al. (2003). "Phase I trial of intranasal and endobronchial administration of a recombinant adeno-associated virus serotype 2 (rAAV2)-CFTR vector in adult cystic fibrosis patients: a two-part clinical study." Hum Gene Ther **14**(11): 1079-1088.

Fondong, V. N. (2013). "Geminivirus protein structure and function." Mol Plant Pathol **14**(6): 635-649.

Fouts, E. T., X. Yu, et al. (1999). "Biochemical and electron microscopic image analysis of the hexameric E1 helicase." J Biol Chem **274**(7): 4447-4458.

Friedman-Einat, M., Z. Grossman, et al. (1997). "Detection of adeno-associated virus type 2 sequences in the human genital tract." J Clin Microbiol **35**(1): 71-78.

Gabriel, R., A. Lombardo, et al. (2011). "An unbiased genome-wide analysis of zinc-finger nuclease specificity." Nat Biotechnol **29**(9): 816-823.

Gai, D., D. Li, et al. (2004). "Insights into the oligomeric states, conformational changes, and helicase activities of SV40 large tumor antigen." J Biol Chem **279**(37): 38952-38959.

Gao, G., L. H. Vandenberghe, et al. (2004). "Clades of Adeno-associated viruses are widely disseminated in human tissues." J Virol **78**(12): 6381-6388.

Gao, G. P., M. R. Alvira, et al. (2002). "Novel adeno-associated viruses from rhesus monkeys as vectors for human gene therapy." Proc Natl Acad Sci U S A **99**(18): 11854-11859.

Geoffroy, M. C. and A. Salvetti (2005). "Helper functions required for wild type and recombinant adeno-associated virus growth." Curr Gene Ther **5**(3): 265-271.

Gerry, H. W., T. J. Kelly, Jr., et al. (1973). "Arrangement of nucleotide sequences in adeno-associated virus DNA." J Mol Biol **79**(2): 207-225.

Ghosh, A. and D. Duan (2007). "Expanding adeno-associated viral vector capacity: a tale of two vectors." Biotechnol Genet Eng Rev **24**: 165-177.

Giraud, C., E. Winocour, et al. (1994). "Site-specific integration by adeno-associated virus is directed by a cellular DNA sequence." Proc Natl Acad Sci U S A **91**(21): 10039-10043.

Girod, A., C. E. Wobus, et al. (2002). "The VP1 capsid protein of adeno-associated virus type 2 is carrying a phospholipase A2 domain required for virus infectivity." J Gen Virol **83**(Pt 5): 973-978.

Gomez-Lorenzo, M. G., M. Valle, et al. (2003). "Large T antigen on the simian virus 40 origin of replication: a 3D snapshot prior to DNA replication." EMBO J **22**(23): 6205-6213.

Gonzalez, F., S. Boue, et al. (2011). "Methods for making induced pluripotent stem cells: reprogramming a la carte." Nat Rev Genet **12**(4): 231-242.

Gonzalez-Prieto, C., L. Agundez, et al. (2013). "HUH site-specific recombinases for targeted modification of the human genome." Trends Biotechnol **31**(5): 305-312.

Gorbalenya, A. E. and E. V. Koonin (1993). "Helicases: amino acid sequence comparisons and structure-function relationships." Current Opinion in Structural Biology **3**(3): 419-429.

Graham, F. L., J. Smiley, et al. (1977). "Characteristics of a human cell line transformed by DNA from human adenovirus type 5." J Gen Virol **36**(1): 59-74.

Green, M. R. and R. G. Roeder (1980). "Transcripts of the adeno-associated virus genome: mapping of the major RNAs." J Virol **36**(1): 79-92.

Grieger, J. C., V. W. Choi, et al. (2006a). "Production and characterization of adeno-associated viral vectors." Nat Protoc **1**(3): 1412-1428.

Grieger, J. C. and R. J. Samulski (2012). "Adeno-associated virus vectorology, manufacturing, and clinical applications." Methods Enzymol **507**: 229-254.

Grieger, J. C., S. Snowdy, et al. (2006b). "Separate basic region motifs within the adeno-associated virus capsid proteins are essential for infectivity and assembly." J Virol **80**(11): 5199-5210.

Grimm, D., A. Kern, et al. (1998). "Novel tools for production and purification of recombinant adenoassociated virus vectors." Hum Gene Ther **9**(18): 2745-2760.

Guilbaud, M., G. Chadeuf, et al. (2008). "Relative influence of the adeno-associated virus (AAV) type 2 p5 element for recombinant AAV vector site-specific integration." J Virol **82**(5): 2590-2593.

Hanson, P. I. and S. W. Whiteheart (2005). "AAA+ proteins: have engine, will work." Nat Rev Mol Cell Biol **6**(7): 519-529.

Hauswirth, W. W. and K. I. Berns (1977). "Origin and termination of adeno-associated virus DNA replication." Virology **78**(2): 488-499.

Heilbronn, R., M. Engstler, et al. (2003). "ssDNA-dependent colocalization of adeno-associated virus Rep and herpes simplex virus ICP8 in nuclear replication domains." Nucleic Acids Res **31**(21): 6206-6213.

Henckaerts, E., N. Dutheil, et al. (2009). "Site-specific integration of adeno-associated virus involves partial duplication of the target locus." Proc Natl Acad Sci U S A **106**(18): 7571-7576.

Henckaerts, E. and R. M. Linden (2010). "Adeno-associated virus: a key to the human genome?" Future Virol **5**(5): 555-574.

Hermonat, P. L. (1994). "Down-regulation of the human c-fos and c-myc proto-oncogene promoters by adeno-associated virus Rep78." Cancer Lett **81**(2): 129-136.

Hermonat, P. L. and R. B. Batchu (1997). "The adeno-associated virus Rep78 major regulatory protein forms multimeric complexes and the domain for this activity is contained within the carboxy-half of the molecule." FEBS Lett **401**(2-3): 180-184.

Hermonat, P. L., M. A. Labow, et al. (1984). "Genetics of adeno-associated virus: isolation and preliminary characterization of adeno-associated virus type 2 mutants." J Virol **51**(2): 329-339.

Hermonat, P. L. and N. Muzyczka (1984). "Use of adeno-associated virus as a mammalian DNA cloning vector: transduction of neomycin resistance into mammalian tissue culture cells." Proc Natl Acad Sci U S A **81**(20): 6466-6470.

Hickman, A. B. and F. Dyda (2005). "Binding and unwinding: SF3 viral helicases." Curr Opin Struct Biol **15**(1): 77-85.

Hickman, A. B., D. R. Ronning, et al. (2002). "Structural unity among viral origin binding proteins: crystal structure of the nuclease domain of adeno-associated virus Rep." Mol Cell **10**(2): 327-337.

Hickman, A. B., D. R. Ronning, et al. (2004). "The nuclease domain of adeno-associated virus rep coordinates replication initiation using two distinct DNA recognition interfaces." Mol Cell **13**(3): 403-414.

High, K. A. and P. Aubourg (2011). "rAAV human trial experience." Methods Mol Biol **807**: 429-457.

Hildinger, M., L. Baldi, et al. (2007). "High-titer, serum-free production of adeno-associated virus vectors by polyethyleneimine-mediated plasmid transfection in mammalian suspension cells." Biotechnol Lett **29**(11): 1713-1721.

Hirt, B. (1967). "Selective extraction of polyoma DNA from infected mouse cell cultures." Journal of molecular biology **26**(2): 365-369.

Hoggan, M. D., N. R. Blacklow, et al. (1966). "Studies of small DNA viruses found in various adenovirus preparations: physical, biological, and immunological characteristics." Proc Natl Acad Sci U S A **55**(6): 1467-1474.

Hoggan, M. D., Thomas, G. F., and Johnson, F. B. (1972). "Continuous carriage of adenovirus-associated virus genome in cell culture in the absence of helper adenovirus." Proceedings of the Fourth Lepetit Colloquium: 41-47.

Hong, G., P. Ward, et al. (1992). "In vitro replication of adeno-associated virus DNA." Proc Natl Acad Sci U S A **89**(10): 4673-4677.

Horer, M., S. Weger, et al. (1995). "Mutational analysis of adeno-associated virus Rep protein-mediated inhibition of heterologous and homologous promoters." J Virol **69**(9): 5485-5496.

Huser, D., A. Gogol-Doring, et al. (2014). "Adeno-associated virus type 2 wild-type and vector-mediated genomic integration profiles of human diploid fibroblasts analyzed by third-generation PacBio DNA sequencing." J Virol **88**(19): 11253-11263.

Huser, D., A. Gogol-Doring, et al. (2010). "Integration preferences of wildtype AAV-2 for consensus rep-binding sites at numerous loci in the human genome." PLoS Pathog **6**(7): e1000985.

Huser, D., S. Weger, et al. (2002). "Kinetics and frequency of adeno-associated virus site-specific integration into human chromosome 19 monitored by quantitative real-time PCR." J Virol **76**(15): 7554-7559.

Ilyina, T. V. and E. V. Koonin (1992). "Conserved sequence motifs in the initiator proteins for rolling circle DNA replication encoded by diverse replicons from eubacteria, eucaryotes and archaebacteria." Nucleic Acids Res **20**(13): 3279-3285.

Im, D. S. and N. Muzyczka (1990). "The AAV origin binding protein Rep68 is an ATP-dependent site-specific endonuclease with DNA helicase activity." Cell **61**(3): 447-457.

Inoue, H., N. Nagata, et al. (2014). "iPS cells: a game changer for future medicine." EMBO J **33**(5): 409-417.

Iyer, L. M., D. D. Leipe, et al. (2004). "Evolutionary history and higher order classification of AAA+ ATPases." J Struct Biol **146**(1-2): 11-31.

James, J. A., A. K. Aggarwal, et al. (2004). "Structure of adeno-associated virus type 2 Rep40-ADP complex: insight into nucleotide recognition and catalysis by superfamily 3 helicases." Proc Natl Acad Sci U S A **101**(34): 12455-12460.

James, J. A., C. R. Escalante, et al. (2003). "Crystal structure of the SF3 helicase from adeno-associated virus type 2." Structure **11**(8): 1025-1035.

Janik, J. E., M. M. Huston, et al. (1989). "Efficient synthesis of adeno-associated virus structural proteins requires both adenovirus DNA binding protein and VA I RNA." Virology **168**(2): 320-329.

Janovitz, T., I. A. Klein, et al. (2013). "High-throughput sequencing reveals principles of adeno-associated virus serotype 2 integration." J Virol **87**(15): 8559-8568.

Jinek, M., K. Chylinski, et al. (2012). "A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity." Science **337**(6096): 816-821.

Johnson, F. B., H. L. Ozer, et al. (1971). "Structural proteins of adenovirus-associated virus type 3." J Virol **8**(6): 860-863.

Johnson, J. S. and R. J. Samulski (2009). "Enhancement of adeno-associated virus infection by mobilizing capsids into and out of the nucleolus." J Virol **83**(6): 2632-2644.

Kay, M. A., C. S. Manno, et al. (2000). "Evidence for gene transfer and expression of factor IX in haemophilia B patients treated with an AAV vector." Nat Genet **24**(3): 257-261.

Kearns, W. G., S. A. Afione, et al. (1996). "Recombinant adeno-associated virus (AAV-CFTR) vectors do not integrate in a site-specific fashion in an immortalized epithelial cell line." Gene Ther **3**(9): 748-755.

Kerr, J. R. and R. M. Linden (2006). Human dependovirus infection. Parvoviruses. J. R. Kerr, S. F. Cotmore, M. E. Bloom, R. M. Linden and C. R. Parrish. London, Hodder Arnold**:** 381-383.

Kim, H. and J. S. Kim (2014). "A guide to genome engineering with programmable nucleases." Nat Rev Genet **15**(5): 321-334.

King, J. A., R. Dubielzig, et al. (2001). "DNA helicase-mediated packaging of adeno-associated virus type 2 genomes into preformed capsids." EMBO J **20**(12): 3282-3291.

Koczot, F. J., B. J. Carter, et al. (1973). "Self-complementarity of terminal sequences within plus or minus strands of adenovirus-associated virus DNA." Proc Natl Acad Sci U S A **70**(1): 215-219.

Kohlbrenner, E., E. Henckaerts, et al. (2012). "Quantification of AAV particle titers by infrared fluorescence scanning of coomassie-stained sodium dodecyl sulfate-polyacrylamide gels." Hum Gene Ther Methods **23**(3): 198-203.

Kotin, R. M., R. M. Linden, et al. (1992). "Characterization of a preferred site on human chromosome 19q for integration of adeno-associated virus DNA by non-homologous recombination." EMBO J **11**(13): 5071-5078.

Kotin, R. M., J. C. Menninger, et al. (1991). "Mapping and direct visualization of a region-specific viral DNA integration site on chromosome 19q13-qter." Genomics **10**(3): 831-834.

Kotin, R. M., M. Siniscalco, et al. (1990). "Site-specific integration by adeno-associated virus." Proc Natl Acad Sci U S A **87**(6): 2211-2215.

Kronenberg, S., B. Bottcher, et al. (2005). "A conformational change in the adeno-associated virus type 2 capsid leads to the exposure of hidden VP1 N termini." J Virol **79**(9): 5296-5303.

Kyostio, S. R. and R. A. Owens (1996). "Identification of mutant adeno-associated virus Rep proteins which are dominant-negative for DNA helicase activity." Biochem Biophys Res Commun **220**(2): 294-299.

Kyostio, S. R., R. A. Owens, et al. (1994). "Analysis of adeno-associated virus (AAV) wild-type and mutant Rep proteins for their abilities to negatively regulate AAV p5 and p19 mRNA levels." J Virol **68**(5): 2947-2957.

Kyostio, S. R., R. S. Wonderling, et al. (1995). "Negative regulation of the adeno-associated virus (AAV) P5 promoter involves both the P5 rep binding site and the consensus ATP-binding motif of the AAV Rep68 protein." J Virol **69**(11): 6787-6796.

Labow, M. A., P. L. Hermonat, et al. (1986). "Positive and negative autoregulation of the adeno-associated virus type 2 genome." J Virol **60**(1): 251-258.

Lackner, D. F. and N. Muzyczka (2002). "Studies of the mechanism of transactivation of the adeno-associated virus p19 promoter by Rep protein." J Virol **76**(16): 8225-8235.

Lamartina, S., G. Ciliberto, et al. (2000a). "Selective cleavage of AAVS1 substrates by the adeno-associated virus type 2 rep68 protein is dependent on topological and sequence constraints." J Virol **74**(19): 8831-8842.

Lamartina, S., E. Sporeno, et al. (2000b). "Characteristics of the adeno-associated virus preintegration site in human chromosome 19: open chromatin conformation and transcription-competent environment." J Virol **74**(16): 7671-7677.

Laughlin, C. A., C. B. Cardellichio, et al. (1986). "Latent infection of KB cells with adeno-associated virus type 2." J Virol **60**(2): 515-524.

Laughlin, C. A., J. D. Tratschin, et al. (1983). "Cloning of infectious adeno-associated virus genomes in bacterial plasmids." Gene **23**(1): 65-73.

Laughlin, C. A., H. Westphal, et al. (1979). "Spliced adenovirus-associated virus RNA." Proc Natl Acad Sci U S A **76**(11): 5567-5571.

Lee, Y. B., C. P. Glover, et al. (2005). "Optimizing regulatable gene expression using adenoviral vectors." Exp Physiol **90**(1): 33-37.

Lefebvre, R. B., S. Riva, et al. (1984). "Conformation takes precedence over sequence in adeno-associated virus DNA replication." Mol Cell Biol **4**(7): 1416-1419.

Lerch, T. F., Q. Xie, et al. (2010). "The structure of adeno-associated virus serotype 3B (AAV-3B): insights into receptor binding and immune evasion." Virology **403**(1): 26-36.

Levy, H. C., V. D. Bowman, et al. (2009). "Heparin binding induces conformational changes in Adeno-associated virus serotype 2." J Struct Biol **165**(3): 146-156.

Li, D., R. Zhao, et al. (2003a). "Structure of the replicative helicase of the oncoprotein SV40 large tumour antigen." Nature **423**(6939): 512-518.

Li, J., R. J. Samulski, et al. (1997). "Role for highly regulated rep gene expression in adeno-associated virus vector production." J Virol **71**(7): 5236-5243.

Li, Z., J. R. Brister, et al. (2003b). "Characterization of the adenoassociated virus Rep protein complex formed on the viral origin of DNA replication." Virology **313**(2): 364-376.

Linden, R. M., P. Ward, et al. (1996a). "Site-specific integration by adeno-associated virus." Proc Natl Acad Sci U S A **93**(21): 11288-11294.

Linden, R. M., E. Winocour, et al. (1996b). "The recombination signals for adeno-associated virus site-specific integration." Proc Natl Acad Sci U S A **93**(15): 7966-7972.

Lombardo, A., D. Cesana, et al. (2011). "Site-specific integration and tailoring of cassette design for sustainable gene transfer." Nat Methods **8**(10): 861-869.

Lusby, E., K. H. Fife, et al. (1980). "Nucleotide sequence of the inverted terminal repetition in adeno-associated virus DNA." J Virol **34**(2): 402-409.

Maggin, J. E., J. A. James, et al. (2012). "The amino acid linker between the endonuclease and helicase domains of adeno-associated virus type 5 Rep plays a critical role in DNA-dependent oligomerization." J Virol **86**(6): 3337-3346.

Mansilla-Soto, J., M. Yoon-Robarts, et al. (2009). "DNA structure modulates the oligomerization properties of the AAV initiator protein Rep68." PLoS Pathog **5**(7): e1000513.

Marcus, C. J., C. A. Laughlin, et al. (1981). "Adeno-associated virus RNA transcription in vivo." Eur J Biochem **121**(1): 147-154.

Matsushita, T., S. Elliger, et al. (1998). "Adeno-associated virus vectors can be efficiently produced without helper virus." Gene Ther **5**(7): 938-945.

McCarty, D. M. (2008). "Self-complementary AAV vectors; advances and applications." Mol Ther **16**(10): 1648-1656.

McCarty, D. M., M. Christensen, et al. (1991). "Sequences required for coordinate induction of adeno-associated virus p19 and p40 promoters by Rep protein." J Virol **65**(6): 2936-2945.

McCarty, D. M., T. H. Ni, et al. (1992). "Analysis of mutations in adeno-associated virus Rep protein in vivo and in vitro." J Virol **66**(7): 4050-4057.

McCarty, D. M., D. J. Pereira, et al. (1994a). "Identification of linear DNA sequences that specifically bind the adeno-associated virus Rep protein." J Virol **68**(8): 4988-4997.

McCarty, D. M., J. H. Ryan, et al. (1994b). "Interaction of the adeno-associated virus Rep protein with a sequence within the A palindrome of the viral terminal repeat." J Virol **68**(8): 4998-5006.

McCarty, D. M., S. M. Young, Jr., et al. (2004). "Integration of adeno-associated virus (AAV) and recombinant AAV vectors." Annu Rev Genet **38**: 819-845.

McLaughlin, S. K., P. Collis, et al. (1988). "Adeno-associated virus general transduction vectors: analysis of proviral structures." J Virol **62**(6): 1963-1973.

Mendelson, E., J. P. Trempe, et al. (1986). "Identification of the trans-acting Rep proteins of adeno-associated virus by antibodies to a synthetic oligopeptide." J Virol **60**(3): 823-832.

Miller, D. G., L. M. Petek, et al. (2004). "Adeno-associated virus vectors integrate at chromosome breakage sites." Nat Genet **36**(7): 767-773.

Miller, E. B., B. Gurda-Whitaker, et al. (2006). "Production, purification and preliminary X-ray crystallographic studies of adeno-associated virus serotype 1." Acta Crystallogr Sect F Struct Biol Cryst Commun **62**(Pt 12): 1271-1274.

Millet, R., N. Jolinon, et al. (2015). "Impact of the MRN Complex on Adeno-Associated Virus Integration and Replication during Coinfection with Herpes Simplex Virus 1." J Virol **89**(13): 6824-6834.

Mori, S., L. Wang, et al. (2004). "Two novel adeno-associated viruses from cynomolgus monkey: pseudotyping characterization of capsid protein." Virology **330**(2): 375-383.

Morita, E., J. Arii, et al. (2012). "Attenuated protein expression vectors for use in siRNA rescue experiments." BioTechniques: 1-5.

Mouw, M. B. and D. J. Pintel (2000). "Adeno-associated virus RNAs appear in a temporal order and their splicing is stimulated during coinfection with adenovirus." J Virol **74**(21): 9878-9888.

Mueller, C. and T. R. Flotte (2008). "Clinical gene therapy using recombinant adeno-associated virus vectors." Gene Ther **15**(11): 858-863.

Muramatsu, S., H. Mizukami, et al. (1996). "Nucleotide sequencing and generation of an infectious clone of adeno-associated virus 3." Virology **221**(1): 208-217.

Musayev, F. N., F. Zarate-Perez, et al. (2015a). "Structural Studies of AAV2 Rep68 Reveal a Partially Structured Linker and Compact Domain Conformation." Biochemistry.

Musayev, F. N., F. Zarate-Perez, et al. (2015b). "Structural Insights into the Assembly of the Adeno-Associated Virus Type 2 Rep68 Protein on the Integration site AAVS1." J Biol Chem.

Myers, M. W. and B. J. Carter (1980). "Assembly of adeno-associated virus." Virology **102**(1): 71-82.

Nam, H. J., B. L. Gurda, et al. (2011). "Structural studies of adeno-associated virus serotype 8 capsid transitions associated with endosomal trafficking." J Virol **85**(22): 11791-11799.

Nam, H. J., M. D. Lane, et al. (2007). "Structure of adeno-associated virus serotype 8, a gene therapy vector." J Virol **81**(22): 12260-12271.

Nash, K., W. Chen, et al. (2008). "Complete in vitro reconstitution of adeno-associated virus DNA replication requires the minichromosome maintenance complex proteins." J Virol **82**(3): 1458-1464.

Needham, P. G., J. M. Casper, et al. (2006). "Adeno-associated virus rep protein-mediated inhibition of transcription of the adenovirus major late promoter in vitro." J Virol **80**(13): 6207-6217.

Ng, R., L. Govindasamy, et al. (2010). "Structural characterization of the dual glycan binding adeno-associated virus serotype 6." J Virol **84**(24): 12945-12957.

Ni, T. H., X. Zhou, et al. (1994). "In vitro replication of adeno-associated virus DNA." J Virol **68**(2): 1128-1138.

Nicolson, S. C. and R. J. Samulski (2014). "Recombinant adeno-associated virus utilizes host cell nuclear import machinery to enter the nucleus." J Virol **88**(8): 4132-4144.

Nonnenmacher, M. and T. Weber (2011). "Adeno-associated virus 2 infection requires endocytosis through the CLIC/GEEC pathway." Cell Host Microbe **10**(6): 563-576.

Nonnenmacher, M. and T. Weber (2012). "Intracellular transport of recombinant adeno-associated virus vectors." Gene Ther **19**(6): 649-658.

Owens, R. A., J. P. Trempe, et al. (1991). "Adeno-associated virus rep proteins produced in insect and mammalian expression systems: wild-type and dominant-negative mutant proteins bind to the viral replication origin." Virology **184**(1): 14-22.

Owens, R. A., M. D. Weitzman, et al. (1993). "Identification of a DNA-binding domain in the amino terminus of adeno-associated virus Rep proteins." J Virol **67**(2): 997-1005.

Padron, E., V. Bowman, et al. (2005). "Structure of adeno-associated virus type 4." J Virol **79**(8): 5047-5058.

Park, J. Y., B. P. Lim, et al. (2006). "Scalable production of adeno-associated virus type 2 vectors via suspension transfection." Biotechnol Bioeng **94**(3): 416-430.

Paruzynski, A., A. Arens, et al. (2010). "Genome-wide high-throughput integrome analyses by nrLAM-PCR and next-generation sequencing." Nat Protoc **5**(8): 1379-1395.

Patel, S. S. and I. Donmez (2006). "Mechanisms of helicases." J Biol Chem **281**(27): 18265-18268.

Patel, S. S., M. Pandey, et al. (2011). "Dynamic coupling between the motors of DNA replication: hexameric helicase, DNA polymerase, and primase." Curr Opin Chem Biol **15**(5): 595-605.

Pereira, D. J., D. M. McCarty, et al. (1997). "The adeno-associated virus (AAV) Rep protein acts as both a repressor and an activator to regulate AAV transcription during a productive infection." J Virol **71**(2): 1079-1088.

Pereira, D. J. and N. Muzyczka (1997a). "The adeno-associated virus type 2 p40 promoter requires a proximal Sp1 interaction and a p19 CArG-like element to facilitate Rep transactivation." J Virol **71**(6): 4300-4309.

Pereira, D. J. and N. Muzyczka (1997b). "The cellular transcription factor SP1 and an unknown cellular protein are required to mediate Rep protein activation of the adeno-associated virus p19 promoter." J Virol **71**(3): 1747-1756.

Petri, K., R. Gabriel, et al. (2015). "Presence of a trs-Like Motif Promotes Rep-Mediated Wild-Type Adeno-Associated Virus Type 2 Integration." J Virol **89**(14): 7428-7432.

Philpott, N. J., J. Gomos, et al. (2002). "A p5 integration efficiency element mediates Rep-dependent integration into AAVS1 at chromosome 19." Proc Natl Acad Sci U S A **99**(19): 12381-12385.

Popa-Wagner, R., M. Porwal, et al. (2012). "Impact of VP1-specific protein sequence motifs on adeno-associated virus type 2 intracellular trafficking and nuclear entry." J Virol **86**(17): 9163-9174.

Qing, K., J. Hansen, et al. (2001). "Adeno-associated virus type 2-mediated gene transfer: role of cellular FKBP52 protein in transgene expression." J Virol **75**(19): 8968-8976.

Qiu, J. and D. J. Pintel (2002). "The adeno-associated virus type 2 Rep protein regulates RNA processing via interaction with the transcription template." Mol Cell Biol **22**(11): 3639-3652.

Qiu, J., Y. Yoto, et al. (2006). Parvovirus RNA processing strategies. Parvoviruses. J. R. Kerr, S. F. Cotmore, M. E. Bloom, R. M. Linden and C. R. Parrish. Lodnon, Hodder Arnold**:** 253-275.

Quesada, O., B. Gurda, et al. (2007). "Production, purification and preliminary X-ray crystallographic studies of adeno-associated virus serotype 7." Acta Crystallogr Sect F Struct Biol Cryst Commun **63**(Pt 12): 1073-1076.

Rabinowitz, J. E., F. Rolling, et al. (2002). "Cross-packaging of a single adeno-associated virus (AAV) type 2 vector genome into multiple AAV serotypes enables transduction with broad specificity." J Virol **76**(2): 791-801.

Recchia, A. and F. Mavilio (2011). "Site-specific integration by the adeno-associated virus rep protein." Curr Gene Ther **11**(5): 399-405.

Rivat, C., G. Santilli, et al. (2012). "Gene therapy for primary immunodeficiencies." Hum Gene Ther **23**(7): 668-675.

Rose, J. A., K. I. Berns, et al. (1969). "Evidence for a single-stranded adenovirus-associated virus genome: formation of a DNA density hybrid on release of viral DNA." Proc Natl Acad Sci U S A **64**(3): 863-869.

Rose, J. A., J. V. Maizel, Jr., et al. (1971). "Structural proteins of adenovirus-associated viruses." J Virol **8**(5): 766-770.

Rutledge, E. A., C. L. Halbert, et al. (1998). "Infectious clones and vectors derived from adeno-associated virus (AAV) serotypes other than AAV type 2." J Virol **72**(1): 309-319.

Ryan, J. H., S. Zolotukhin, et al. (1996). "Sequence requirements for binding of Rep68 to the adeno-associated virus terminal repeats." J Virol **70**(3): 1542-1553.

Sadelain, M., E. P. Papapetrou, et al. (2012). "Safe harbours for the integration of new DNA in the human genome." Nat Rev Cancer **12**(1): 51-58.

Salganik, M., B. Venkatakrishnan, et al. (2012). "Evidence for pH-dependent protease activity in the adeno-associated virus capsid." J Virol **86**(21): 11877-11885.

Samulski, R. J., K. I. Berns, et al. (1982). "Cloning of adeno-associated virus into pBR322: rescue of intact virus from the recombinant plasmid in human cells." Proc Natl Acad Sci U S A **79**(6): 2077-2081.

Samulski, R. J., L. S. Chang, et al. (1987). "A recombinant plasmid from which an infectious adeno-associated virus genome can be excised in vitro and its use to study viral replication." J Virol **61**(10): 3096-3101.

Samulski, R. J., L. S. Chang, et al. (1989). "Helper-free stocks of recombinant adeno-associated viruses: normal integration does not require viral gene expression." J Virol **63**(9): 3822-3828.

Samulski, R. J., A. Srivastava, et al. (1983). "Rescue of adeno-associated virus from recombinant plasmids: gene correction within the terminal repeats of AAV." Cell **33**(1): 135-143.

Samulski, R. J., X. Zhu, et al. (1991). "Targeted integration of adeno-associated virus (AAV) into human chromosome 19." EMBO J **10**(12): 3941-3950.

Schlehofer, J. R., M. Ehrbar, et al. (1986). "Vaccinia virus, herpes simplex virus, and carcinogens induce DNA amplification in a human cell line and support replication of a helpervirus dependent parvovirus." Virology **152**(1): 110-117.

Schmidt, M., L. Govindasamy, et al. (2008a). "Molecular characterization of the heparin-dependent transduction domain on the capsid of a novel adeno-associated virus isolate, AAV(VR-942)." J Virol **82**(17): 8911-8916.

Schmidt, M., E. Grot, et al. (2006). "Identification and characterization of novel adeno-associated virus isolates in ATCC virus stocks." J Virol **80**(10): 5082-5085.

Schmidt, M., K. Schwarzwaelder, et al. (2007). "High-resolution insertion-site analysis by linear amplification-mediated PCR (LAM-PCR)." Nat Methods **4**(12): 1051-1057.

Schmidt, M., A. Voutetakis, et al. (2008b). "Adeno-associated virus type 12 (AAV12): a novel AAV serotype with sialic acid- and heparan sulfate proteoglycan-independent transduction activity." J Virol **82**(3): 1399-1406.

Schmittgen, T. D. and K. J. Livak (2008). "Analyzing real-time PCR data by the comparative C(T) method." Nat Protoc **3**(6): 1101-1108.

Schnepp, B. C., K. R. Clark, et al. (2003). "Genetic fate of recombinant adeno-associated virus vector genomes in muscle." J Virol **77**(6): 3495-3504.

Schnepp, B. C., R. L. Jensen, et al. (2005). "Characterization of adeno-associated virus genomes isolated from human tissues." J Virol **79**(23): 14793-14803.

Schnepp, B. C., R. L. Jensen, et al. (2009). "Infectious molecular clones of adeno-associated virus isolated directly from human tissues." J Virol **83**(3): 1456-1464.

Schuck, S. and A. Stenlund (2011). "Mechanistic analysis of local ori melting and helicase assembly by the papillomavirus E1 protein." Mol Cell **43**(5): 776-787.

Schultz, B. R. and J. S. Chamberlain (2008). "Recombinant adeno-associated virus transduction and integration." Mol Ther **16**(7): 1189-1199.

Schwartz, R. A., J. A. Palacios, et al. (2007). "The Mre11/Rad50/Nbs1 complex limits adeno-associated virus transduction and replication." J Virol **81**(23): 12936-12945.

Sedman, J. and A. Stenlund (1998). "The papillomavirus E1 protein forms a DNA-dependent hexameric complex with ATPase and DNA helicase activities." J Virol **72**(8): 6893-6897.

Shi, Y., E. Seto, et al. (1991). "Transcriptional repression by YY1, a human GLI-Kruppel-related protein, and relief of repression by adenovirus E1A protein." Cell **67**(2): 377-388.

Singleton, M. R., M. S. Dillingham, et al. (2007). "Structure and mechanism of helicases and nucleic acid translocases." Annu Rev Biochem **76**: 23-50.

Slanina, H., S. Weger, et al. (2006). "Role of the herpes simplex virus helicase-primase complex during adeno-associated virus DNA replication." J Virol **80**(11): 5241-5250.

Smith, R. H. and R. M. Kotin (1998). "The Rep52 gene product of adeno-associated virus is a DNA helicase with 3'-to-5' polarity." J Virol **72**(6): 4874-4881.

Smith, R. H. and R. M. Kotin (2000). "An adeno-associated virus (AAV) initiator protein, Rep78, catalyzes the cleavage and ligation of single-stranded AAV ori DNA." J Virol **74**(7): 3122-3129.

Smith, R. H., A. J. Spano, et al. (1997). "The Rep78 gene product of adeno-associated virus (AAV) self-associates to form a hexameric complex in the presence of AAV ori sequences." J Virol **71**(6): 4461-4471.

Snyder, R. O., D. S. Im, et al. (1990). "Evidence for covalent attachment of the adeno-associated virus (AAV) rep protein to the ends of the AAV genome." J Virol **64**(12): 6204-6213.

Snyder, R. O., D. S. Im, et al. (1993). "Features of the adeno-associated virus origin involved in substrate recognition by the viral Rep protein." J Virol **67**(10): 6096-6104.

Somers, A., J. C. Jean, et al. (2010). "Generation of transgene-free lung disease-specific human induced pluripotent stem cells using a single excisable lentiviral stem cell cassette." Stem Cells **28**(10): 1728-1740.

Sonntag, F., S. Bleker, et al. (2006). "Adeno-associated virus type 2 capsids with externalized VP1/VP2 trafficking domains are generated prior to passage through the cytoplasm and are maintained until uncoating occurs in the nucleus." J Virol **80**(22): 11040-11054.

Sonntag, F., K. Schmidt, et al. (2010). "A viral assembly factor promotes AAV2 capsid formation in the nucleolus." Proc Natl Acad Sci U S A **107**(22): 10220-10225.

Srivastava, A., E. W. Lusby, et al. (1983). "Nucleotide sequence and organization of the adeno-associated virus 2 genome." J Virol **45**(2): 555-564.

Stahnke, S., K. Lux, et al. (2011). "Intrinsic phospholipase A2 activity of adeno-associated virus is involved in endosomal escape of incoming particles." Virology **409**(1): 77-83.

Sternberg, S. H. and J. A. Doudna (2015). "Expanding the Biologist's Toolkit with CRISPR-Cas9." Mol Cell **58**(4): 568-574.

Stracker, T. H., G. D. Cassell, et al. (2004). "The Rep protein of adeno-associated virus type 2 interacts with single-stranded DNA-binding proteins that enhance viral replication." J Virol **78**(1): 441-453.

Straus, S. E., E. D. Sebring, et al. (1976). "Concatemers of alternating plus and minus strands are intermediates in adenovirus-associated virus DNA synthesis." Proc Natl Acad Sci U S A **73**(3): 742-746.

Streck, C. J., P. V. Dickson, et al. (2006). "Antitumor efficacy of AAV-mediated systemic delivery of interferon-beta." Cancer Gene Ther **13**(1): 99-106.

Summerford, C. and R. J. Samulski (1998). "Membrane-associated heparan sulfate proteoglycan is a receptor for adeno-associated virus type 2 virions." J Virol **72**(2): 1438-1445.

Surosky, R. T., M. Urabe, et al. (1997). "Adeno-associated virus Rep proteins target DNA sequences to a unique locus in the human genome." J Virol **71**(10): 7951-7959.

Takahashi, K. and S. Yamanaka (2006). "Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors." Cell **126**(4): 663-676.

Tattersall, P. (2006). The evolution of parvovirus taxonomy. Parvoviruses. J. R. Kerr, S. F. Cotmore, M. E. Bloom, R. M. Linden and C. R. Parrish. London, Hodder Arnold**:** 5-15.

Tattersall, P. and D. C. Ward (1976). "Rolling hairpin model for replication of parvovirus and linear chromosomal DNA." Nature **263**(5573): 106-109.

Tezak, Z., K. Nagaraju, et al. (2000). "Adeno-associated virus in normal and myositis human skeletal muscle." Neurology **55**(12): 1913-1917.

Titolo, S., A. Pelletier, et al. (2000). "Identification of domains of the human papillomavirus type 11 E1 helicase involved in oligomerization and binding to the viral origin." J Virol **74**(16): 7349-7361.

Tiyaboonchai, A., H. Mac, et al. (2014). "Utilization of the AAVS1 safe harbor locus for hematopoietic specific transgene expression and gene knockdown in human ES cells." Stem Cell Res **12**(3): 630-637.

Tratschin, J. D., I. L. Miller, et al. (1984a). "Genetic analysis of adeno-associated virus: properties of deletion mutants constructed in vitro and evidence for an adeno-associated virus replication function." J Virol **51**(3): 611-619.

Tratschin, J. D., M. H. West, et al. (1984b). "A human parvovirus, adeno-associated virus, as a eucaryotic vector: transient expression and encapsidation of the procaryotic gene for chloramphenicol acetyltransferase." Mol Cell Biol **4**(10): 2072-2081.

Trempe, J. P. and B. J. Carter (1988). "Regulation of adeno-associated virus gene expression in 293 cells: control of mRNA abundance and translation." J Virol **62**(1): 68-74.

Urabe, M., Y. Hasumi, et al. (1999). "Charged-to-alanine scanning mutagenesis of the N-terminal half of adeno-associated virus type 2 Rep78 protein." J Virol **73**(4): 2682-2693.

Urcelay, E., P. Ward, et al. (1995). "Asymmetric replication in vitro from a human sequence element is dependent on adeno-associated virus Rep protein." J Virol **69**(4): 2038-2046.

Valle, M., C. Gruss, et al. (2000). "Large T-antigen double hexamers imaged at the simian virus 40 origin of replication." Mol Cell Biol **20**(1): 34-41.

Valle, N., L. Riolobos, et al. (2006). Synthesis, post-translational modification and trafficking of the parvovirus structural peptides. Parvoviruses. J. R. Kerr, S. F. Cotmore, M. E. Bloom, R. M. Linden and C. R. Parrish. London, Hodder Arnold**:** 291-304.

Vandenberghe, L. H., J. M. Wilson, et al. (2009). "Tailoring the AAV vector capsid for gene therapy." Gene Ther **16**(3): 311-319.

Vandenberghe, L. H., R. Xiao, et al. (2010). "Efficient serotype-dependent release of functional vector into the culture medium during adeno-associated virus manufacturing." Hum Gene Ther **21**(10): 1251-1257.

Virag, T., S. Cecchini, et al. (2009). "Producing recombinant adeno-associated virus in foster cells: overcoming production limitations using a baculovirus-insect cell expression strategy." Hum Gene Ther **20**(8): 807-817.

Virgin, H. W. (2014). "The virome in mammalian physiology and disease." Cell **157**(1): 142-150.

Vogel, R., M. Seyffert, et al. (2013). "Viral and Cellular Components of AAV2 Replication Compartments." Open Virol J **7**: 98-120.

Walker, S. L., R. S. Wonderling, et al. (1997a). "Mutational analysis of the adeno-associated virus Rep68 protein: identification of critical residues necessary for site-specific endonuclease activity." J Virol **71**(4): 2722-2730.

Walker, S. L., R. S. Wonderling, et al. (1997b). "Mutational analysis of the adeno-associated virus type 2 Rep68 protein helicase motifs." J Virol **71**(9): 6996-7004.

Walsh, C. E., J. M. Liu, et al. (1992). "Regulated high level expression of a human gamma-globin gene introduced into erythroid cells by an adeno-associated virus vector." Proc Natl Acad Sci U S A **89**(15): 7257-7261.

Walters, R. W., M. Agbandje-McKenna, et al. (2004). "Structure of adeno-associated virus serotype 5." J Virol **78**(7): 3361-3371.

Walz, C., A. Deprez, et al. (1997). "Interaction of human papillomavirus type 16 and adeno-associated virus type 2 co-infecting human cervical epithelium." J Gen Virol **78 (Pt 6)**: 1441-1452.

Ward, P. (2006). Replication of adeno-associated virus DNA. Parvoviruses. J. R. Kerr, S. F. Cotmore, M. E. Bloom, R. M. Linden and C. R. Parrish. London, Hodder Arnold**:** 189-211.

Ward, P. and K. I. Berns (1996). "In vitro replication of adeno-associated virus DNA: enhancement by extracts from adenovirus-infected HeLa cells." J Virol **70**(7): 4495-4501.

Ward, P., F. B. Dean, et al. (1998). "Role of the adenovirus DNA-binding protein in in vitro adeno-associated virus DNA replication." J Virol **72**(1): 420-427.

Ward, P., M. Falkenberg, et al. (2001). "Rep-dependent initiation of adeno-associated virus type 2 DNA replication by a herpes simplex virus type 1 replication complex in a reconstituted system." J Virol **75**(21): 10250-10258.

Weger, S., A. Wistuba, et al. (1997). "Control of adeno-associated virus type 2 cap gene expression: relative influence of helper virus, terminal repeats, and Rep proteins." J Virol **71**(11): 8437-8447.

Weindler, F. W. and R. Heilbronn (1991). "A subset of herpes simplex virus replication genes provides helper functions for productive adeno-associated virus replication." J Virol **65**(5): 2476-2483.

Weisshart, K., S. Friedl, et al. (2004). "Partial proteolysis of simian virus 40 T antigen reveals intramolecular contacts between domains and conformation changes upon hexamer assembly." J Biol Chem **279**(37): 38943-38951.

Weisshart, K., P. Taneja, et al. (1999). "Two regions of simian virus 40 T antigen determine cooperativity of double-hexamer assembly on the viral origin of DNA replication and promote hexamer interactions during bidirectional origin DNA unwinding." J Virol **73**(3): 2201-2211.

Weitzman, M. D., S. R. Kyostio, et al. (1996). "Interaction of wild-type and mutant adeno-associated virus (AAV) Rep proteins on AAV hairpin DNA." J Virol **70**(4): 2440-2448.

Weitzman, M. D., S. R. Kyostio, et al. (1994). "Adeno-associated virus (AAV) Rep proteins mediate complex formation between AAV DNA and its integration site in human DNA." Proc Natl Acad Sci U S A **91**(13): 5808-5812.

Williams, D. A. and A. J. Thrasher (2014). "Concise review: lessons learned from clinical trials of gene therapy in monogenic immunodeficiency diseases." Stem Cells Transl Med **3**(5): 636-642.

Wistuba, A., A. Kern, et al. (1997). "Subcellular compartmentalization of adeno-associated virus type 2 assembly." J Virol **71**(2): 1341-1352.

Wonderling, R. S. and R. A. Owens (1997). "Binding sites for adeno-associated virus Rep proteins within the human genome." J Virol **71**(3): 2528-2534.

Wright, J. F. (2009). "Transient transfection methods for clinical adeno-associated viral vector production." Hum Gene Ther **20**(7): 698-706.

Xiao, P. J. and R. J. Samulski (2012). "Cytoplasmic trafficking, endosomal escape, and perinuclear accumulation of adeno-associated virus type 2 particles are facilitated by microtubule network." J Virol **86**(19): 10462-10473.

Xiao, X., J. Li, et al. (1998). "Production of high-titer recombinant adeno-associated virus vectors in the absence of helper adenovirus." J Virol **72**(3): 2224-2232.

Xie, Q., W. Bu, et al. (2002). "The atomic structure of adeno-associated virus (AAV-2), a vector for human gene therapy." Proc Natl Acad Sci U S A **99**(16): 10405-10410.

Yakobson, B., T. A. Hrynko, et al. (1989). "Replication of adeno-associated virus in cells irradiated with UV light at 254 nm." J Virol **63**(3): 1023-1030.

Yakobson, B., T. Koch, et al. (1987). "Replication of adeno-associated virus in synchronized cells without the addition of a helper virus." J Virol **61**(4): 972-981.

Yalkinoglu, A. O., R. Heilbronn, et al. (1988). "DNA amplification of adeno-associated virus as a response to cellular genotoxic stress." Cancer Res **48**(11): 3123-3129.

Yamanaka, S. (2007). "Strategies and new developments in the generation of patient-specific pluripotent stem cells." Cell Stem Cell **1**(1): 39-49.

Yan, Z., R. Zak, et al. (2002). "Ubiquitination of both adeno-associated virus type 2 and 5 capsid proteins affects the transduction efficiency of recombinant vectors." J Virol **76**(5): 2043-2053.

Yang, L., M. H. Soonpaa, et al. (2008). "Human cardiovascular progenitor cells develop from a KDR+ embryonic-stem-cell-derived population." Nature **453**(7194): 524-528.

Yang, Q., A. Kadam, et al. (1992). "Mutational analysis of the adeno-associated virus rep gene." J Virol **66**(10): 6058-6069.

Yla-Herttuala, S. (2012). "Endgame: glybera finally recommended for approval as the first gene therapy drug in the European union." Molecular therapy : the journal of the American Society of Gene Therapy **20**(10): 1831-1832.

Yoon, M., D. H. Smith, et al. (2001). "Amino-terminal domain exchange redirects origin-specific interactions of adeno-associated virus rep78 in vitro." J Virol **75**(7): 3230-3239.

Yoon-Robarts, M., A. G. Blouin, et al. (2004). "Residues within the B' motif are critical for DNA binding by the superfamily 3 helicase Rep40 of adeno-associated virus type 2." J Biol Chem **279**(48): 50472-50481.

Young, L., J. Sung, et al. (2010). "Detection of Mycoplasma in cell cultures." Nat Protoc **5**(5): 929-934.

Young, S. M., Jr. and R. J. Samulski (2001). "Adeno-associated virus (AAV) site-specific recombination does not require a Rep-dependent origin of replication within the AAV terminal repeat." Proc Natl Acad Sci U S A **98**(24): 13525-13530.

Zarate-Perez, F., M. Bardelli, et al. (2012). "The interdomain linker of AAV-2 Rep68 is an integral part of its oligomerization domain: role of a conserved SF3 helicase residue in oligomerization." PLoS Pathog **8**(6): e1002764.

Zarate-Perez, F., J. Mansilla-Soto, et al. (2013). "Oligomeric properties of adeno-associated virus Rep68 reflect its multifunctionality." J Virol **87**(2): 1232-1241.

Zeltner, N., E. Kohlbrenner, et al. (2010). "Near-perfect infectivity of wild-type AAV as benchmark for infectivity of recombinant AAV vectors." Gene Ther **17**(7): 872-879.

Zhang, F., S. I. Thornhill, et al. (2007). "Lentiviral vectors containing an enhancer-less ubiquitously acting chromatin opening element (UCOE) provide highly reproducible and stable transgene expression in hematopoietic cells." Blood **110**(5): 1448-1457.

Zhang, H., J. Xie, et al. (2009). "Adenovirus-adeno-associated virus hybrid for large-scale recombinant adeno-associated virus production." Hum Gene Ther **20**(9): 922-929.

Zhong, L., B. Li, et al. (2008). "Tyrosine-phosphorylation of AAV2 vectors and its consequences on viral intracellular trafficking and transgene expression." Virology **381**(2): 194-202.

Zhou, X., I. Zolotukhin, et al. (1999). "Biochemical characterization of adeno-associated virus rep68 DNA helicase and ATPase activities." J Virol **73**(2): 1580-1590.

Zincarelli, C., S. Soltys, et al. (2008). "Analysis of AAV serotypes 1-9 mediated gene expression and tropism in mice after systemic injection." Mol Ther **16**(6): 1073-1080.

Zolotukhin, S., M. Potter, et al. (1996). "A "humanized" green fluorescent protein cDNA adapted for high-level expression in mammalian cells." J Virol **70**(7): 4646-4654.

ANNEX 1: SUPPLEMENTARY INFORMATION

TABLE 1: PLASMIDS

Plasmid	Content	Cloning strategy	Resistance
AAV			
pAV2	Infectious AAV plasmids		Amp
pSub201	Infectious AAV plasmids		Amp
pMB2	pAV2 based, C151S Rep mutation	SfiI+HindIII from pMB1	Amp
pMB7	pMB2 based, Y224A Rep mutation	SfiI+HindIII from pET15b-Rep68-C151S-Y224A	Amp
pMB8	pMB2 based, Y224P Rep mutation	SfiI+HindIII from pET15b-Rep68-C151S-Y224P	Amp
pMB9	pMB2 based, Y224F Rep mutation	SfiI+HindIII from pET15b-Rep68-C151S-Y224F	Amp
pMB11	pMB2 based, Rep mutations L193A and V196A	SfiI+HindIII from pMB10	Amp
pMB27	pMB2 based, Rep mutations Y224A and I251A	SfiI+HindIII from pMB28	Amp
pMB31	pMB2 based, Rep mutations I251A	SfiI+HindIII from pMB29	Amp
pMB32	pMB2 based, Rep mutations K340H	SfiI+HindIII from pMB30	Amp
Helpers			
mini-pDG	Rep and CAP only		Amp
HGTI	Adeno helper functions		Amp
pXX6	Adeno helper functions		Amp
pAAVDJ	AAV2 Rep and DJ Cap		Amp
pMB1	mini-pDG based, C151 Rep mutation	mutagenesis on mini-pDG	Amp
pMB4	pMB1 based, Y224A Rep mutation	SfiI+HindIII from pET15b-Rep68-C151S-Y224A	Amp
pMB5	pMB1 based, Y224P Rep mutation	SfiI+HindIII from pET15b-Rep68-C151S-Y224P	Amp
pMB6	pMB1 based, Y224F Rep mutation	SfiI+HindIII from pET15b-Rep68-C151S-Y224F	Amp
pMB10	pMB1 based, Rep mutations L193A and V196A	mutagenesis on pMB1	Amp
pMB28	pMB1 based, Rep mutations Y224A and I251A	mutagenesis on pMB1	Amp
pMB29	pMB1 based, Rep mutation I251A	mutagenesis on pMB1	Amp
pMB30	pMB1 based, Rep mutation 340H	BamHI+BstEII from pND232	Amp
rAAV			
pTRUF11	CAG-GFP-pk-NEO between ITRs		Amp
pMB12	pSUB201-EF1a-CG (no WPRE, no polyA tail)	EF1a-CG from SINLV-CG with XbaI, SbfI (upstream) and AvrII, XhoI, NotI, NheI (downstream)overhangs by PCR cloned into pSUB201 XbaI sites inside the ITRs. The second XbaI site was lost because it was ligated with the compatible NheI site in the insert.	Amp
pMB13	pSUB201-EF1a-CG	WPRE and SV40 (late) polyA from pLA100 (LA/AA) with AvrII (upstream) andLoxP, XhoI, NotI (downstream) overhangs by PCR, cloned into pMB12 using AvrII and NotI.	Amp
pMB14	pSUB201-EF1a-CG-loxP-CMV-GFP-loxP	CMV-GFP-pA(bGH) from CMV-GFP plasmid with XhoI (upstream) and LoxP-NotI overhangs (downstream), cloned into pMB13 using XhoI and NotI.	Amp

Plasmid	Content	Cloning strategy	Resistance
Rep			
pND226	CMV-Rep68 Y156F M225G		Kana
pMB21	CMV-Rep68 M225G	From pND226 after mutagenesis to revert F156 to Y	Kana
pMB22	CMV-Rep68 C151S M225G	From pMB21 after mutagenesis C151S	Kana
pMB23	CMV-Rep68 C151S M225G Y224A	mutagenesis on pMB22	Kana
pMB24	CMV-Rep68 C151S M225G Y224F	mutagenesis on pMB22	Kana
pMB25	CMV-Rep68 C151S L193A V196A M225G	mutagenesis on pMB22	Kana
pMB26	CMV-Rep68 C151S M225G Y224A I251A	BamHI+BstEII from pMB28, into pMB23	Kana
pMB33	CMV-Rep68 C151S M225G I251A	BamHI+BstEII from pND29	Kana
pMB34	CMV-Rep68 C151S M225G K340H	BamHI+BstEII from pND232	Kana

TABLE 2: PRIMERS

Name		Sequence (5'->3')	Target
Primers for mutagenesis			
MB1	fw	CAAGGTGGTGGATGAGTCTTACATCCCCAATTAC	mutagenesis of Rep C151 to S151
MB2	rv	GTAATTGGGGATGTAAGACTCATCCACCACCTTG	mutagenesis of Rep C151 to S152
MB18	fw	CGGTTGGTGGCGCAGCATGCCACGCACGCTTCGCAGACGCAGGAGC	Rep mutagenesis: L193A + V196A
MB19	rv	GCTCCTGCGTCTGCGAAGCGTGCGTGGCATGCTGCGCCACCAACCG	Rep mutagenesis: L193A + V196A
MB65_Y224A_fw	fw	CAAAAACTTCAGCCAGGgcCATGGAGCTGGTCGGG	mutagenesis Y224A (wt) in Rep wt
MB66_Y224A_rv	rv	CCCGACCAGCTCCATGgcCCTGGCTGAAGTTTTTG	mutagenesis Y224A (wt) in Rep wt
MB67_Y224F_fw	fw	CAAAAACTTCAGCCAGGTtCATGGAGCTGGTCGGG	mutagenesis Y224F (wt) in Rep wt
MB68_Y224F_rv	rv	CCCGACCAGCTCCATGaACCTGGCTGAAGTTTTTG	mutagenesis Y224F (wt) in Rep wt
MB69_Y224A_fw_M225G	fw	CAAAAACTTCAGCCAGGgcCggaGAGCTGGTCGGG	mutagenesis Y224A (wt) in Rep M225G
MB70_Y224A_rv_M225G	rv	CCCGACCAGCTCtccGgcCCTGGCTGAAGTTTTTG	mutagenesis Y224A (wt) in Rep M225G
MB71_Y224F_fw_M225G	fw	CAAAAACTTCAGCCAGGTtCggaGAGCTGGTCGGG	mutagenesis Y224F (wt) in Rep M225G
MB72_Y224F_rv_M225G	rv	CCCGACCAGCTCtccGaACCTGGCTGAAGTTTTTG	mutagenesis Y224F (wt) in Rep M225G
MB77_251fw	rv	GACCAGGCCTCATACGCCTCCTTCAATGCGGC	mutagenesis I251A in Rep wt
MB78_251rv	fw	GCCGCATTGAAGGAGGCGTATGAGGCCTGGTC	mutagenesis I251A in Rep wt
Primers for cloning			
MB12_Step1fw	fw	ccaaTCTAGACGTTGTCCCTGCAGGttatcgattggctccggtgcccgt	primer for 1st step (EF1a-CG) of cloning of AAV-CG-GFP vector
MB13_Step1rv	rv	ccaaGCTAGCGGGCCGCcaacCTCGAGcaacCCTAGGtaggtttcaggctttagggtgtaacatgggg	primer for 1st step (EF1a-CG) of cloning of AAV-CG-GFP vector
MB14_Step2fw	fw	ccaaCCTAGGCGCCgatccgtcgataatcaacctctggattacaa	primer for 2nd step (WPRE-pA) of cloning of AAV-CG-GFP vector
MB15_Step2rv	rv	ccaaCTCGAGataacttcgtatagcatacattatacgaagttatGGTACCctgggggatcttcgatgctagacgat	primer for 2nd step (WPRE-pA) of cloning of AAV-CG-GFP vector
MB16_Step3fw	fw	ccaaCTCGAGcgatgtacgggccagatatacgcgttgaca	primer for 3rd step (CMV-GFP) of cloning of AAV-CG-GFP vector
MB17_Step3rv	rv	ccaaGCGGCCGCataacttcgtatagcatacattatacgaagttatATGCATagaagccatagagcccaccgcatcccc	primer for 3rd step (CMV-GFP) of cloning of AAV-CG-GFP vector
MB81_Rep68_fw	fw	CCAAtctagaATGCCGGGGTTTTACGAGATTGTG	amplification of Rep68 from CMV-Rep68, XbaI and NotI overhangs
MB82_Rep68_rv	rv	CCAAgcggccgcTCAGAGAGAGTGTCCTCGAGCCAATC	amplification of Rep68 from CMV-Rep68, XbaI and NotI overhangs

Name		Sequence (5'->3')	Target
Primers for sequencing			
MB3	rv	GTCAGGCTCATAATCTTTCCCGCA	sequencing of Rep
MB20	fw	cttcacaaactgagtgaatcccagctag	sequencing of IL2RG in AAV vectors
MB21	rv	tgctccaaacagtggttcaagaatctg	sequencing of IL2RG in AAV vectors
MB22	fw	TCCTGGTTGCTGTCTCTTTATG	sequencing of pMB14 (WPRE)
MB23	fw	CGATGTACGGGCCAGATATAC	sequencing of pMB14
MB24	fw	GTAAACGGCCACAAGTTCAG	sequencing of pMB14 (GFP)
MB25	rv	AGTTCCCGTGGTATTCAGTAAC	sequencing of pMB14 (IL2RG)
ND140	fw	GTTTCCTGAGTCAGATTCGCG	sequencing of Rep
qPCR primers			
Cap1	fw	TTCTCAGATGCTGCGTACCGGAAA	wt AAV quantification
Cap2	rv	TCTGCCATTGAGGTGGTACTTGGT	wt AAV quantification
CycloFW	fw	TGCTGGACCCAACACAAATG	wt AAV quantification
CycloRV	rv	TGCCATCCAACCACTCAGTCT	wt AAV quantification
MB85_p5fw	rv	AACAAGGTGGTGGATGAGT	taqman qPCR on AAV cDNA
MB86_p5rv	fw	CGTTTACGCTCCGTGAGATT	taqman qPCR on AAV cDNA
MB87_p19fw	fw	TCACCAAGCAGGAAGTCAAAG	taqman qPCR on AAV cDNA
MB88_p19rv	rv	CCCGTTTGGGCTCACTTATATC	taqman qPCR on AAV cDNA
MB89_p40fw	fw	GGAAGCAAGGCTCAGAGAAA	taqman qPCR on AAV cDNA
MB90_p40rv	rv	CCTCTCTGGAGGTTGGTAGATA	taqman qPCR on AAV cDNA
CGfw	fw	CGCCACCATGTTGAAGCCATCATT	titration of CG vectors
CGrv	rv	TCAGCTGTGGTGTCTTCATTCCCA	titration of CG vectors
qPCR probes			
probe p5	fw	FAM-ACTGTTCCATATTAGTCCACGCCCAC-TAM	FAM-TAMRA probe to be used with primers MB85 and MB86
probe p19	fw	FAM_ACGTGGTTGAGGTGGAGCATGAAT-TAM	FAM-TAMRA probe to be used with primers MB87 and MB88
probe p40	fw	FAM-AGGAAATCAGGACAACCAATCCCGT-TAM	FAM-TAMRA probe to be used with primers MB89 and MB90
Primers for SB probes			
Rep1	fw	AACTGGACCAATGAGAACTTTCC	amplification of Rep probe
Rep2	rv	AAAAAGTCTTTGACTTCCTGCTT	amplification of Rep probe
MB4-ampfw	fw	AATCAGTGAGGCACCTATCTCAGC	amplification of ampicillin probe
MB5-amprv	rv	AACTCGGTCGCCGCATACACTATT	amplification of ampicillin probe

TABLE 3: ANTIBODIES

Primary antibodies

Antibody	Species	Source (Clone)	Dilution	Application
α-HSP90	rabbit, polyclonal	Santa Cruz Biotechnology (H114)	1:10000	WB
α-Rep	mouse, monoclonal	Progen (303.9)	1:200 / 1:100	WB, IF
α-VP1,2,3	mouse, monoclonal	ARP (B1)	1:200	WB
α-GFP	mouse, monoclonal	Roche (7.1/13.1)	1:5000	WB
α-γChain	mouse, monoclonal	Santa Cruz Biotechnology (E7)	1:500	WB
α-γChain-BV650	rat, monoclonal	BD Biosciences (TUGh4)	1:500	FACS
α-PSTAT5-Alexa647	mouse, polyclonal	BD Biosciences (PY694)	1:500	FACS
α-nanog	rabbit, polyclonal	Abcam	1:100	IF
α-SSEA-4	mouse, monoclonal	Millipore (MC-813-70)	1:100	IF
α-Tra1-60	mouse, monoclonal	Millipore (TRA-1-60)	1:100	IF

Secondary antibodies

Antibody	Conjugation	Source	Dilution	Application
α-mouse IgG	Dylight 680	Cell Signalling Technology	1:5000	WB, LI-COR Odyssey
α-rabbit IgG	Dylight 800	Cell Signalling Technology	1:5000	WB, LI-COR Odyssey
Goat α-mouse IgG	HRP	Biorad	1:10000	WB, ImageQuant
Goat α-rabbit IgG	HRP	Biorad	1:10000	WB, ImageQuant
Goat α-rabbit IgG	Alexa Fluor® 488	Invitrogen Life Technologies	1:5000	IF
Donkey α-mouse IgG	Alexa Fluor® 594	Invitrogen Life Technologies	1:5000	IF
Donkey α-mouse IgG & IgM	Alexa Fluor® 488	Invitrogen Life Technologies	1:5000	IF

TABLE 4: REP MUTANTS REPORTED IN THE LITERATURE

Table legend:

- ○ Mutant performs similarly to wt in this assay
- ◌ Mutant has an intermediate phenotype
- ● Mutant is completely non-functional

Different papers often use different assays to test for the same function, so caution should be used when comparing results from different papers. This is particularly the case for papers that studied Rep-mediated integration and transcriptional regulation. Please refer to the indicated references for details on the assays used.

Mutant	Nuclear localisation	Replication	RBS binding	dsTRS nicking	ssTRS nicking	Helicase	ATPase	Transcriptional regulation	Oligomerisation	Integration	Dominant negative	References
Point mutants												
P2A	○	○						○				Kleinschmidt (Virology, 1995)
Y5F			○	○	○							Davis (JVI, 2000)
E6A			○	●						●		Urabe (JVI, 1999)
E83C				●	●							Yoon-Robarts (JBC, 2003)
K10A			◍	●						●		Urabe (JVI, 1999)
D14A			○	○						○		Urabe (JVI, 1999)
D14C				○	○							Yoon-Robarts (JBC, 2003)
D16A			○	●						●		Urabe (JVI, 1999)
D16C				○	○							Yoon-Robarts (JBC, 2003)
E17A			○	○						○		Urabe (JVI, 1999)
H18A			○	○						○		Urabe (JVI, 1999)
D24			○	○						○		Urabe (JVI, 1999)
D24C				○	○							Yoon-Robarts (JBC, 2003)
E32A-K33A			○	○						○		Urabe (JVI, 1999)
E34A			○	○						○		Urabe (JVI, 1999)
E36A			○	○						○		Urabe (JVI, 1999)
D40A			○	○						○		Urabe (JVI, 1999)
D40A-D42A-D44A			●	●	●	◍	◍		◍			Davis (JVI, 1999), Davis (JVI, 2000)
D42A			○	○	○					●		Urabe (JVI, 1999)
D44A			○	○						○		Urabe (JVI, 1999)
E49A			○	○						○		Urabe (JVI, 1999)
E57A-K58A			○	○						○		Urabe (JVI, 1999)
R61A-D62A			○	○						○		Urabe (JVI, 1999)
E66A			◍	●						◍		Urabe (JVI, 1999)
E66A-K72A			◍	◍		◍	◍		◍			Davis (JVI, 1999)
R68A			○	○						○		Urabe (JVI, 1999)
R69A			◍	●						●		Urabe (JVI, 1999)

Mutant	Nuclear localisation	Replication	RBS binding	dsTRS nicking	ssTRS nicking	Helicase	ATPase	Transcriptional regulation	Oligomerisation	Integration	Dominant negative	References
K72A			○	●						●		Urabe (JVI, 1999)
E75A			○	○						○		Urabe (JVI, 1999)
E83A			○	●						●		Urabe (JVI, 1999)
E83C			○	◐	◐	○						Yoon-Robarts (JBC, 2003)
E83A-K84A-E86A			◐	●	●	○	◐		◐		○	Davis (JVI, 1999), Davis (JVI, 2000)
K84A			○	●						◐		Urabe (JVI, 1999)
E86A			○	○						○		Urabe (JVI, 1999)
Y88F					○							Davis (JVI, 2000)
H90A			○	○						○		Urabe (JVI, 1999)
H92A			○	○						○		Urabe (JVI, 1999)
E96A			○	●						◐		Urabe (JVI, 1999)
K101A			◐	●						●		Urabe (JVI, 1999)
R107A			●	●					◐	●		Urabe (JVI, 1999), Mansilla-Soto (PPat, 2009), Hickman (Mol Cell, 2004)
R113A			◐	●						●		Urabe (JVI, 1999)
E114A			○	○						○		Urabe (JVI, 1999)
E114A-K115A			◐	◐		◐	◐		◐			Davis (JVI, 1999)
K115A			○	○						○		Urabe (JVI, 1999)
R119A			○	○						○		Urabe (JVI, 1999)
Y121F			○	●	○	●	◐				●	Davis (JVI, 2000), Walker (JVI, 1997a)
R122A			○	○						○		Urabe (JVI, 1999)
E125A			○	○						○		Urabe (JVI, 1999)
K136A			●	●						●		Urabe (JVI, 1999)
R138A			●	●						●		Urabe (JVI, 1999)
K146A			○	●						●		Urabe (JVI, 1999)
K146A-D149A-E150A			◐	●	●	○	○		◐			Davis (JVI, 1999), Davis (JVI, 2000)
D149A			○	○						○		Urabe (JVI, 1999)
E150A			○	○						●		Urabe (JVI, 1999)
C151S	○	○	○	○		○	○	○	○		○	Zarate-Perez (JVI, 2013)

Mutant	Nuclear localisation	Replication	RBS binding	dsTRS nicking	ssTRS nicking	Helicase	ATPase	Transcriptional regulation	Oligomerisation	Integration	Dominant negative	References
C151S-C405S		●										unpublished
Y152F					◍							Davis (JVI, 2000)
Y156F				●	●							Davis (JVI, 2000)
Y152F-Y156F				●	●							Davis (JVI, 2000)
K160A			○	●						●		Urabe (JVI, 1999)
L165G									◍			Smith (JVI, 1997)
E164A			○	○						●		Urabe (JVI, 1999)
E164C			○	◍	◍	○						Yoon-Robarts (JBC, 2003)
E173A			○	○						○		Urabe (JVI, 1999)
Y175F			○		○							Davis (JVI, 2000)
L180T									○			Smith (JVI, 1997)
T183A	○	○						○				Kleinschmidt (Virology, 1995)
E184A			○	○						○		Urabe (JVI, 1999)
E184A-K186A			○	○		○	◍			○		Davis (JVI, 1999)
R185A			◍	●						●		Urabe (JVI, 1999)
K186A-R187A			○	○						○		Urabe (JVI, 1999)
H192A			○	○						○		Urabe (JVI, 1999)
L193A-V196A			○	○						○		Musayev (JBC, 2015)
H195A			○	○						○		Urabe (JVI, 1999)
E201A			○	○						○		Urabe (JVI, 1999)
E201A-K204A-E205A			○	○		○	◍		○			Davis (JVI, 1999)
K204A-E205A			○	○						○		Urabe (JVI, 1999)
D212A			○	○						○		Urabe (JVI, 1999)
R217A			○	●						●		Urabe (JVI, 1999)
K219A			○	●						●		Urabe (JVI, 1999)
S221G	○	●						○				Kleinschmidt (Virology, 1995)
R223A			○	○						●		Urabe (JVI, 1999)
Y224A	○	●	●	●		○		●	●			Zarate-Perez (PPat, 2012)

Mutant	Nuclear localisation	Replication	RBS binding	dsTRS nicking	ssTRS nicking	Helicase	ATPase	Transcriptional regulation	Oligomerisation	Integration	Dominant negative	References
Y224F			●	●		●	●					Walker (JVI, 1997a)
Y224F	○	◐	○	◐		○		◐	◐	●		submitted
Y224A-I251A	○	●	●	●		○		●	●			submitted
M225G	○	○						○				Kyostio (JVI, 1995), Kleinschmidt (Virology, 1995)
E226A			○	○						○		Urabe (JVI, 1999)
D233A			○	○						○		Urabe (JVI, 1999)
K234A			○	●						●		Urabe (JVI, 1999)
E239A			◐	○						◐		Urabe (JVI, 1999)
K240A			○	●						●		Urabe (JVI, 1999)
W242L/S		●	●	●				◐				McCarty (JVI, 1992), Pereira (JVI, 1997)
Y250F			○	○		○	○					Walker (JVI, 1997a)
I251A	○	●	●	●		○		●	●			submitted
Y283F			○	○		○	○					Walker (JVI, 1997a)
Y299F			○	○		○	○					Walker (JVI, 1997a)
Y307F			◐	●		●	◐					Walker (JVI, 1997a)
Y311F			◐	●		○	◐					Walker (JVI, 1997a)
G334A			○		○	●	●					Davis (JVI, 2000), Walker (JVI, 1997b)
G339A			○		○	●	●					Davis (JVI, 2000), Walker (JVI, 1997b)
K340H	○	●	○	●	○	●	●	●	○	○	●	Chejanowsky (1990), Owens (1991), Kyostio (1994), Kleinschmidt (Virology, 1995), Kyostio and Owens (1996), Smith (JVI, 1997), Urabe (JVI, 1999), Davis (JVI, 1999), Davis (JVI, 2000)
T341A			○		○	●	●					Davis (JVI, 2000), Walker (JVI, 1997b)
T341I-N342Y		●	○	●				●				McCarty (JVI, 1992), Pereira (JVI, 1997)
Y354F			○	○		○	○					Walker (JVI, 1997a)
P365A		●	○	◐				◐				McCarty (JVI, 1992), Pereira (JVI, 1997)
P365T		○	○	○				○				McCarty (JVI, 1992), Pereira (JVI, 1997)
E378A-E379A-K381A			◐		○	●	●		◐			Davis (JVI, 1999), Davis (JVI, 2000)
E379A			◐	●		●	◐					Walker (JVI, 1997b)
E379K/Q		●	◐	●				◐				McCarty (JVI, 1992), Pereira (JVI, 1997)

Mutant	Nuclear localisation	Replication	RBS binding	dsTRS nicking	ssTRS nicking	Helicase	ATPase	Transcriptional regulation	Oligomerisation	Integration	Dominant negative	References
E391A			◐	●		●	●					Walker (JVI, 1997b)
K391I/T		●	○	●				◐				McCarty (JVI, 1992), Pereira (JVI, 1997)
I393A			◐	●		●	◐					Walker (JVI, 1997b)
G395A			○	○		○	○					Walker (JVI, 1997b)
G396A			○	○		○	○					Walker (JVI, 1997b)
D402A-K404A-K406A			○		○	●	◐			○	●	Davis (JVI, 1999), Davis (JVI, 2000)
K404A			○			●	○					Walker (JVI, 1997b), Yoon-Robarts (JBC, 2002), Mansilla-Soto (PPat, 2009)
K404I/T		●	○	●					●			McCarty (JVI, 1992), Pereira (JVI, 1997)
K406A						◐	○		○			Yoon-Robarts (JBC, 2002), Mansilla-Soto (PPat, 2009)
P415H		●	●	●				◐				McCarty (JVI, 1992), Pereira (JVI, 1997)
V416A			○	○		◐	◐					Walker (JVI, 1997b)
I417A			○		○	●	●					Walker (JVI, 1997b)
V418S			○		○	●	●					Walker (JVI, 1997b), Davis (JVI, 2000)
T419A			◐			◐	●					Walker (JVI, 1997b)
D429E			◐	●		●	●					Walker (JVI, 1997a)
D443A-K447A			◐	●	○	●	●		◐		○	Davis (JVI, 1999), Davis (JVI, 2000)
R444A						●	●					Yoon-Robarts (JBC, 2002), James (Structure, 2003)
D455A-D457A			◐	●	○	●	◐		◐		○	Davis (JVI, 1999), Davis (JVI, 2000)
K463A-E465A-K467A-D468A			◐	●	○	●	●		◐		○	Davis (JVI, 1999), Davis (JVI, 2000)
E465D			○	●		○	◐					Walker (JVI, 1997a)
K474A-D475A			◐	●		●	◐		◐			Davis (JVI, 1999)
E479A-E481A-E483A			◐	●		●	●		◐		○	Davis (JVI, 1999)
K492A-K493A-R493A	●	●	●	●	●	●	●	●	●	●		Cassel (Virology, 2004)
K492A-K493A-R493A-K506A-R507A-R509A	●	◐	○	○								Cassel (Virology, 2004)
P505H/L		○	○	○				○				McCarty (JVI, 1992), Pereira (JVI, 1997)
K492A-K493A-R493A	●	◐	○	○								Cassel (Virology, 2004)

Mutant	Nuclear localisation	Replication	RBS binding	dsTRS nicking	ssTRS nicking	Helicase	ATPase	Transcriptional regulation	Oligomerisation	Integration	Dominant negative	References
Deletions												
1-163										◉		Smith (JVI, 1997)
1-171	○	●						○				Kleinschmidt (Virology, 1995)
1-171+530-621+ R529S	○	●						○				Kleinschmidt (Virology, 1995)
1-224 (R52..)		●						◉				Kleinschmidt (Virology, 1995)
1-242										●		Smith (JVI, 1997)
1-369										●		Smith (JVI, 1997)
1-389										●		Smith (JVI, 1997)
1-521										●		Smith (JVI, 1997)
25-56		●	●					●				Yang (JVI, 1993), Yang (JVI, 1992)
57-76		●	●					●				Yang (JVI, 1993), Yang (JVI, 1992)
62-87		●	○					●				Yang (JVI, 1993), Yang (JVI, 1992)
88-113		●	●					●				Yang (JVI, 1993), Yang (JVI, 1992)
114-122		○	○					○				Yang (JVI, 1993), Yang (JVI, 1992)
125-133		●	●					●				Yang (JVI, 1993), Yang (JVI, 1992)
134-143		●	●	●				●				McCarty (JVI, 1992), Pereira (JVI, 1997)
134-164		●	●					●				Yang (JVI, 1993), Yang (JVI, 1992)
144-153		●	●	●				●				McCarty (JVI, 1992), Pereira (JVI, 1997)
151-188										◉		Smith (JVI, 1997)
151-188+334-347										●		Smith (JVI, 1997)
154-163		●	●	●				●				McCarty (JVI, 1992), Pereira (JVI, 1997)
164-173		●	●	●				●				McCarty (JVI, 1992), Pereira (JVI, 1997)
165-217		●	●					●				Yang (JVI, 1993), Yang (JVI, 1992)
165-226		●	●					●				Yang (JVI, 1993), Yang (JVI, 1992)
174-183		●	●	●				●				McCarty (JVI, 1992), Pereira (JVI, 1997)
183-187, P2A	○	◉						◉				Kleinschmidt (Virology, 1995)
197-212, P2A	○	●						○				Kleinschmidt (Virology, 1995)
218-226		●	●					●				Yang (JVI, 1993), Yang (JVI, 1992)

Mutant	Nuclear localisation	Replication	RBS binding	dsTRS nicking	ssTRS nicking	Helicase	ATPase	Transcriptional regulation	Oligomerisation	Integration	Dominant negative	References
218-259		●	●					●				Yang (JVI, 1993), Yang (JVI, 1992)
221-224, P2A	○	●						○				Kleinschmidt (Virology, 1995)
225			◐	○				◐				Davis (JVI, 2000), Kyostio (JVI, 1994)
227-259		●	●			●		●			●	Yang (JVI, 1993), Yang (JVI, 1992), Kyostio (Biochem. Biophys. Res. Commun., 1996)
244-621										●		Smith (JVI, 1997)
245-621, Q244L	●	●						●				Kleinschmidt (Virology, 1995)
257-320		●	○					●				Yang (JVI, 1993), Yang (JVI, 1992)
287-320		●	○					●				Yang (JVI, 1993), Yang (JVI, 1992)
321-346		●	○					●				Yang (JVI, 1993), Yang (JVI, 1992)
323-621											●	Smith (JVI, 1997)
334-347											◐	Smith (JVI, 1997)
347-409		●	●					●				Yang (JVI, 1993), Yang (JVI, 1992)
372-621	●	●						●		◐		Kleinschmidt (Virology, 1995), Smith (JVI, 1997)
401-455		●	○					●				Yang (JVI, 1993), Yang (JVI, 1992)
456-520	●	●	○					●				Yang (JVI, 1993), Yang (JVI, 1992)
461-470, K460L		●	●		●			●				McCarty (JVI, 1992), Pereira (JVI, 1997)
471-480, F470L		●	●		●			●				McCarty (JVI, 1992), Pereira (JVI, 1997)
481-490		●	●		●			●				McCarty (JVI, 1992), Pereira (JVI, 1997)
482-556	●	●	○					●				Yang (JVI, 1993), Yang (JVI, 1992)
485-621											●	Smith (JVI, 1997)
491-500		○	○	○				○				McCarty (JVI, 1992), Pereira (JVI, 1997)
501-510		○	○	○				○				McCarty (JVI, 1992), Pereira (JVI, 1997)
523-621	◐	◐						●				Kleinschmidt (Virology, 1995)
530-621, R529S	○	○						○				Kleinschmidt (Virology, 1995)
549-621											●	Smith (JVI, 1997)

ANNEX 2: PUBLICATIONS DURING THE PHD

Structural Studies of AAV2 Rep68 Reveal a Partially Structured Linker and Compact Domain Conformation.

Musayev FN, Zarate-Perez F, Bardelli M, Bishop C, Saniev EF, Linden RM, Henckaerts E, Escalante CR.

Biochemistry. 2015 Sep 29;54(38):5907-19. doi: 10.1021/acs.biochem.5b00610. Epub 2015 Sep 14

Oligomeric properties of adeno-associated virus Rep68 reflect its multifunctionality.

Zarate-Perez F, Mansilla-Soto J, Bardelli M, Burgner JW 2nd, Villamil-Jarauta M, Kekilli D, Samso M, Linden RM, Escalante CR.

J Virol. 2013 Jan;87(2):1232-41. doi: 10.1128/JVI.02441-12. Epub 2012 Nov 14.

The interdomain linker of AAV-2 Rep68 is an integral part of its oligomerization domain: role of a conserved SF3 helicase residue in oligomerization.

Zarate-Perez F, Bardelli M, Burgner JW 2nd, Villamil-Jarauta M, Das K, Kekilli D, Mansilla-Soto J, Linden RM, Escalante CR.

PLoS Pathog. 2012;8(6):e1002764. doi: 10.1371/journal.ppat.1002764. Epub 2012 Jun 14.

Manuscript submitted:

Identification of a functionally relevant AAV Rep oligomerisation interface.

Bardelli M, Zarate-Perez F, Linden RM, Escalante CR, Henckaerts E

BIOCHEMISTRY
including biophysical chemistry & molecular biology

Article

pubs.acs.org/biochemistry

Structural Studies of AAV2 Rep68 Reveal a Partially Structured Linker and Compact Domain Conformation

Faik N. Musayev,[†,@] Francisco Zarate-Perez,[‡,@] Martino Bardelli,[§] Clayton Bishop,[‡] Emil F. Saniev,[‡] R. Michael Linden,[§,∥] Els Henckaerts,[§] and Carlos R. Escalante*[,‡]

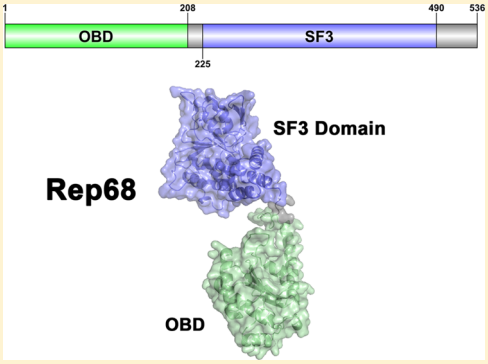
[†]Department of Medicinal Chemistry, School of Pharmacy, [‡]Department of Physiology and Biophysics, School of Medicine, Virginia Commonwealth University, Richmond, Virginia 23298, United States

[§]Department of Infectious Diseases, King's College London, London SE1 9RT, United Kingdom

[∥]UCL Gene Therapy Consortium, UCL Cancer Institute, University College London, London WC1E 6DD, United Kingdom

Supporting Information

ABSTRACT: Adeno-associated virus (AAV) nonstructural proteins Rep78 and Rep68 carry out all DNA transactions that regulate the AAV life cycle. They share two multifunctional domains: an N-terminal origin binding/nicking domain (OBD) from the HUH superfamily and a SF3 helicase domain. A short linker of ∼20 amino acids that is critical for oligomerization and function connects the two domains. Although X-ray structures of the AAV5 OBD and AAV2 helicase domains have been determined, information about the full-length protein and linker conformation is not known. This article presents the solution structure of AAV2 Rep68 using small-angle X-ray scattering (SAXS). We first determined the X-ray structures of the minimal AAV2 Rep68 OBD and of the OBD with the linker region. These X-ray structures reveal novel features that include a long C-terminal α-helix that protrudes from the core of the protein at a 45° angle and a partially structured linker. SAXS studies corroborate that the linker is not extended, and we show that a proline residue in the linker is critical for Rep68 oligomerization and function. SAXS-based rigid-body modeling of Rep68 confirms these observations, showing a compact arrangement of the two domains in which they acquire a conformation that positions key residues in all domains on one face of the protein, poised to interact with DNA.



The nonstructural Rep proteins from adeno-associated virus (AAV) are multifunctional proteins with specialized domains equipped to handle the complex interactions with DNA during the AAV life cycle.[1,2] AAV has a single-stranded DNA genome of ∼4.7 kb containing two major open reading frames (ORFs) flanked by inverted terminal repeats (ITRs). The ITRs form T-shaped hairpin structures and contain the cis-regulatory functions required for replication, transcriptional regulation, and possibly site-specific integration.[3−9] The stem of the hairpin forms a double-stranded region containing a Rep binding site (RBS) made up of several 5′-GCTC-3′ repeats.[10] A terminal resolution site (trs) located upstream of the RBS is the site of a strand- and site-specific endonuclease reaction required to complete the replication of the AAV genome.[3,11,12] Despite its limited genome size, AAV generates eight different polypeptide chains. The right ORF regulated by the P_{40} promoter produces three capsid proteins (VP1−3) and an assembly activating protein (AAP).[13−18] The left ORF uses two different promoters, and alternative splicing, to generate four nonstructural proteins: two large Rep proteins (Rep78 and Rep68) transcribed from the p5 promoter and two small Rep proteins (Rep52 and Rep40) regulated through the p19 promoter.[14,19−21] Thus, Rep40 is equivalent to the helicase domain of Rep68. Most of the biochemical activities required for transcriptional regulation,[22,23] DNA replication[24,25] and site-specific integration[26−28] are carried out by the large Rep78/Rep68 proteins, whereas the small Reps are thought to be important for DNA packaging into preformed capsids.[29,30]

The wide spectrum of functions performed by Rep78/Rep68 is a direct reflection of the multifunctionality of their domains. The N-terminal origin binding domain (OBD) displays site- and strand-specific endonuclease activity and recognizes the GCTC repeats.[24,31−34] Structurally, the OBD is a member of the HUH endonuclease family specialized in the cleavage and rejoining of single-stranded DNA substrates (ssDNA), which occurs during transposition, bacterial conjugation, rolling-circle replication of bacterial DNA and bacteriophages, and in the replication of small eukaryotic viruses.[35] This family is characterized by two signature motifs that participate in the nuclease reaction: the HUH motif, which consists of two histidine (H) residues separated by a large hydrophobic residue (U), and the Y motif, with either one or two tyrosine (Y) residues. Structurally, the common feature of HUH endonucleases is a five-stranded antiparallel β-sheet surrounded by α-helices, where the HUH motif is found in one of the strands.

Received: June 3, 2015
Revised: August 20, 2015
Published: August 28, 2015

DOI: 10.1021/acs.biochem.5b00610
Biochemistry 2015, 54, 5907−5919

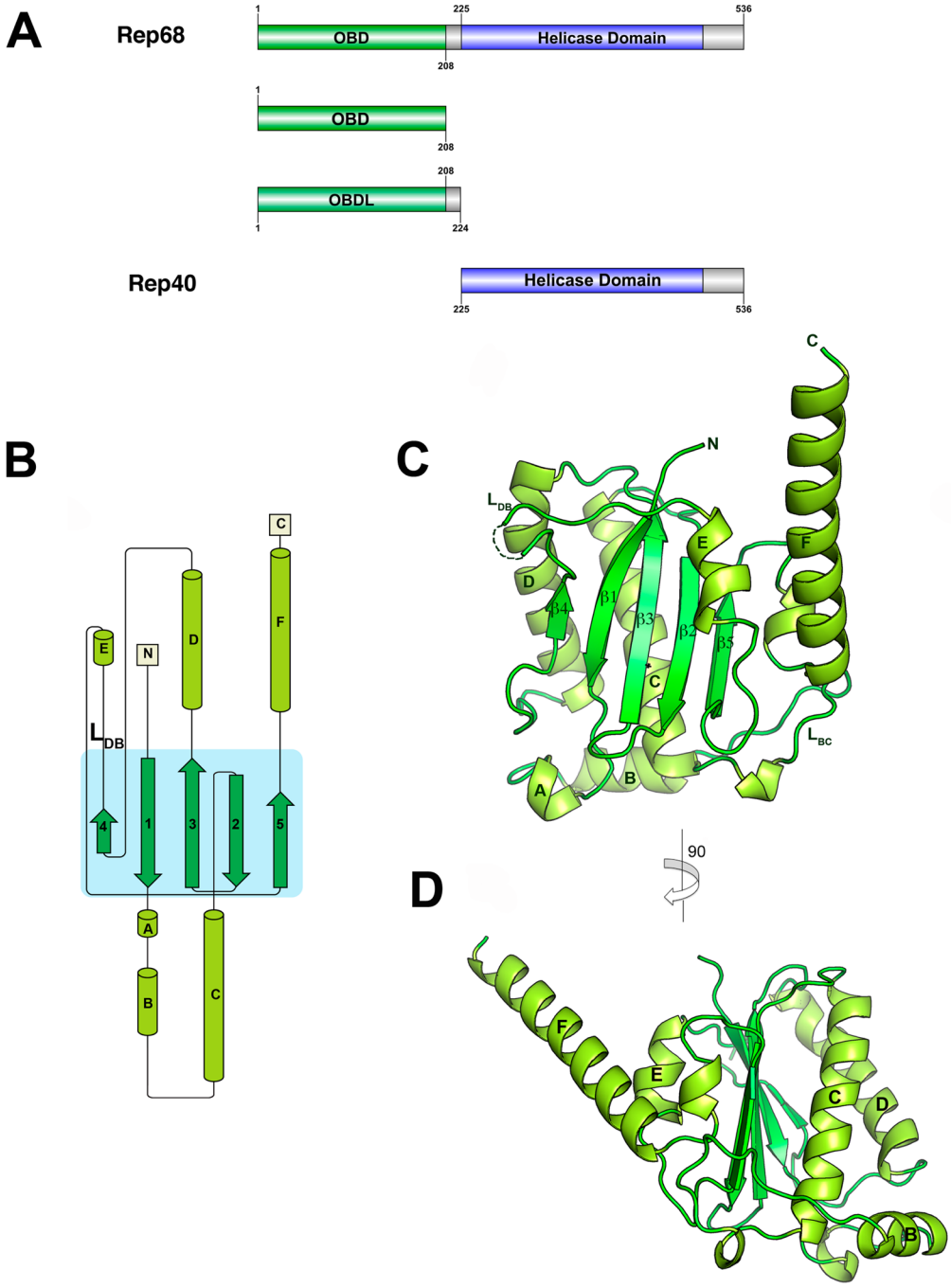


Figure 1. Structure of AAV2 OBD. (A) Domain structure of AAV2 Rep68 protein: OBD is shown in green, Rep40 (SF3 helicase domain), in blue, and linker and C-terminal tail, in gray. The Rep40 protein spans residues 225−536 of Rep68. (B) Topology diagram of AAV2 OBD. (C) Ribbon diagram of the OBD structure. The α-helices are light green, and β-strands are dark green. Secondary structure elements are labeled with α-helices A−F, and β-sheets are numbered 1−5. The DNA binding loop L_{DB} with missing residues in the structure is shown as a dotted line. (D) A view of the structure rotated by 90° clockwise. α-Helix F is shown protruding from the core of the structure at an angle of almost 45°.

The position of the Y motif can vary: in the relaxase subgroup, it is located N-terminally, whereas it is found at the C-terminus in the Rep subclass.[36−38]

The helicase domain found in all Rep proteins is a representative of the SF3 family helicase with 3′ to 5′ unwinding activity.[38−43] The ATPase core is a modified version of the AAA^+ domain with additional features that include an N-terminal helical bundle and a β-hairpin that is involved in DNA binding during translocation.[38] Consequently, the large Rep proteins can interact with DNA in two different modes: one is mediated by the OBD and recognizes DNA in a sequence-specific manner, whereas the SF3 helicase domain interacts with DNA nonspecifically.[44] Furthermore, the large Rep proteins show a complex and dynamic oligomerization behavior, which may add an additional level of regulation to Rep interactions with DNA.[45] How the two domains cooperate with each other in order to interact with DNA and how they are arranged in the context of Rep78/Rep68 is not known. Moreover, studies have established that the linker region plays a critical role in the oligomerization and function of AAV Rep

DOI: 10.1021/acs.biochem.5b00610
Biochemistry 2015, 54, 5907−5919

proteins; therefore, it is likely that this region acquires a particular secondary structure and may not simply be extended.[46,47] To answer these questions, we performed X-ray crystallography and small angle X-ray scattering (SAXS) studies on the AAV2 Rep68 functional domains and full-length protein. Our results show that the OBD has a long extended helix that includes part of the linker, which is not extended, but is tilted in a way that brings the two domains into a compact configuration. SAXS studies on a monomeric version of full-length Rep68 confirm this observation and show for the first time the overall domain architecture of a full-length AAV Rep protein.

MATERIALS AND METHODS

Protein Expression and Purification. The DNA region encoding amino acids 1−208 (OBD) and 1−224 (OBD plus linker) from adeno-associated virus type 2 (AAV2) (GeneBank protein_id = AF043303.1) was cloned into pET15b (Novagen) using restriction sites *Nde*I and *Xho*I. The residue C151 was mutated to serine, as it was found to produce disulfide bonds and inhibit crystallization, and Rep-C151S was shown to be fully functional in supporting the AAV life cycle.[45] The OBD constructs were overexpressed in *Escherichia coli* strain BL21 pLysS at 37 °C in Luria−Bertani (LB) broth. IPTG (isopropyl-*β*-D-thiogalactopyranoside) was added to a final concentration of 1 mM when an OD of 0.6 was reached. Cells were harvested after 5 h and stored at −80 °C. The cell pellets were resuspended in binding buffer (20 mM Tris-HCl, 500 mM NaCl, 10 mM imidazole, 10% glycerol, 1 mM TCEP, pH 7.9) and lysed by sonication. The OBD was purified with a Ni-NTA column (Qiagen) using step gradients of 10 and 30 mM imidazole to wash nonspecific proteins binding to the column and then eluted with 100 mM imidazole. Protein was loaded onto a HiLoad desalting column (GE) to change into thrombin buffer (25 mM Tris-HCl, 200 mM NaCl, 10% glycerol pH 8.0). His-tag was cut by addition of thrombin (1 unit/mg) and removed by passing through a Ni-NTA column. The untagged OBD was collected from the flow through, concentrated, and further purified by gel filtration on a HiLoad 16/60 Superdex 75 column (GE Healthcare) previously equilibrated with gel filtration (GF) buffer (25 mM Tris-HCl, 200 mM NaCl, 1 mM TCEP, pH 7.5). The protein was concentrated to ∼40 mg/mL using Millipore Centricon (10 kDa cutoff). AAV2 Rep40 was purified as described elsewhere.[38] AAV-2 Rep68$_{wt}$ (1−536) and Rep68Y224Δ were expressed in *E. coli* strain BL21 pLysS at 18 °C as described in earlier reports.[46] In brief, histidine-tagged Rep68 was purified in a Ni-NTA column, and after PreScission protease cleavage of the His-tag, the protein was purified on a HiLoad Superdex 200 16/60 column (GE Healthcare). OBDL was purified using the same procedure.

Crystallization, X-ray Data Collection, and Structure Determination. Crystallization was carried out using the hanging-drop method with commercially available screening kits at 4 °C. Crystals grew after 2 to 3 days in 50 mM cacodylate, pH 6.5, 80 mM sodium acetate, 15 mM magnesium acetate, and 8−10% isopropanol. Crystals were cryoprotected in reservoir buffer and supplemented with 20% MPD before flash freezing them in liquid nitrogen. The crystals diffracted to 2.3 Å and belonged to space group $P2_12_12$ with unit cell dimensions a = 186.4 Å, b = 154.4 Å, c = 38.8 Å. Diffraction data was collected at the National Synchrotron Light Source (NSLS) at Brookhaven National Laboratory beamline X6a. The data were processed with the program HKL2000,[48] and the structure was solved by molecular replacement using the program PHENIX. We used the structure of the AAV5 OBD as a search model (PDB ID: 1M55). Model building was carried out using PHENIX,[49] and manual building, using the program COOT.[50] OBDL data was collected using our X-ray home source that consists of a Rigaku Micromax 007 X-ray generator and a Raxis IV^{+2} area detector.

Table 1. Data Collection and Refinement Statistics

	OBD (1−208)	OBDL (1−224)
space group	$P2_12_12$	$P2_1$
Cell Dimensions		
a, b, c (Å)	186.4, 154.4, 38.8	75.6, 178.7, 130.4
α, β, γ (deg)	90, 90, 90	90, 91.7, 90
wavelength (Å)	0.9792	1.54
resolution (Å)	30−2.30 (2.34−2.30)	30−2.6 (2.69−2.6)
no. of measured	566 473	426 030
no. of unique	50 960	105 074
data coverage (%)[a]	99.8 (100)	99.1 (98.9)
R_{merge}(%)[a,b]	0.099 (0.402)	0.075 (0.311)
I/σ[a]	17.7 (8.6)	12.4 (4.1)
refinement statistics		
resolution range	30.0−2.3	29.8−2.6
reflections	50 960	105 007
R_{work} (%)[c]	22.5	19.1
R_{free} (%)[d]	24.2	23.8
non-hydrogen atoms	5024	17677
protein	4704	17493
metal	2	11
water	314	173
average B-factors (Å^2)	49.8	59.1
protein	50.0	59.2
metal	49.2	88.4
water	46.4	47.3
RMS Deviations		
bonds (Å)	0.02	0.003
angles (Å)	1.4	1.53
Ramachandran Plot Quality		
most favored (%)	99.0	97.0
additional allowed (%)	1.0	3.0
generously allowed (%)	0	0.0
disallowed (%)	0	0

[a]Values for the outermost shells are given in parentheses. [b]$R_{merge} = \sum|I - \langle I\rangle|/\sum I$, where I is the integrated intensity of a given reflection. [c]$R_{work} = \sum[||F_o| - |F_c||\sum|F_o|$. [d]For R_{free} calculations, 5% of data was excluded from refinement.

Sedimentation Velocity. Sedimentation velocity experiments were carried out using a Beckman Optima XL-I analytical ultracentrifuge (Beckman Coulter Inc.) equipped with an eight-position AN-60Ti rotor. Rep protein samples were loaded in the cells, using, in all cases, the GF buffer. Samples in double sector cells were centrifuged at 25 000 rpm. In all experiments, temperature was kept at 20 °C. Sedimentation profiles were recorded using UV absorption (280 nm) and interference scanning optics. For analysis of the data, the program Sedfit was used to calculate sedimentation coefficient distribution profiles using the Lamm equation.[51]

Small-Angle X-ray Scattering (SAXS). Data were collected at three different concentrations that produce a single homogeneous population, as determined by sedimentation velocity studies. Synchrotron SAXS measurements were performed at

DOI: 10.1021/acs.biochem.5b00610
Biochemistry 2015, 54, 5907−5919

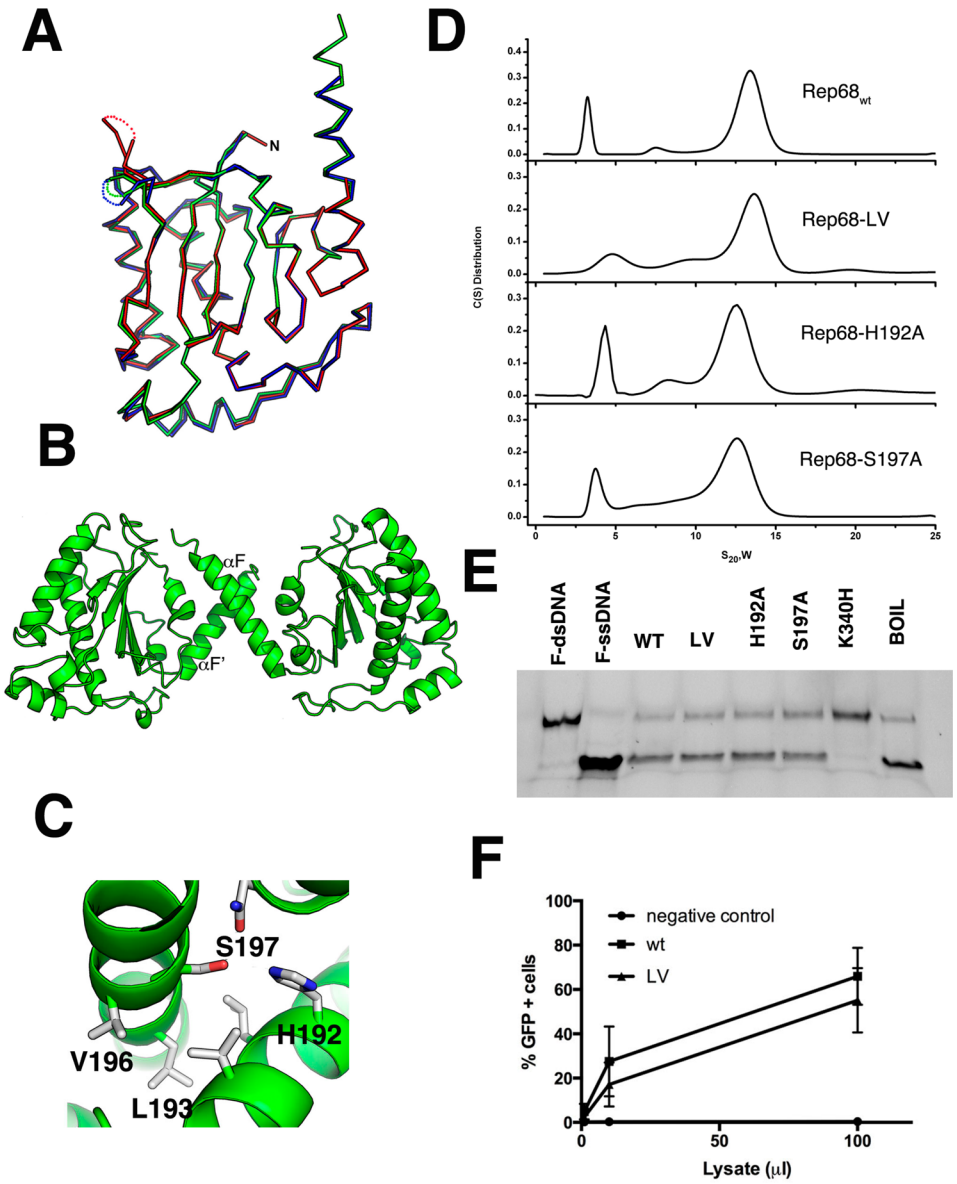


Figure 2. Superposition of OBD molecules in the asymmetric unit. (A) Ribbon diagram of the superposition of the three molecules in the asymmetric unit. Molecule A is shown in green; molecule B, blue; and molecule C, red. The three loops L_{DB} have different conformations and are shown as dotted connections. (B) Ribbon diagram of the dimer in the asymmetric unit formed by molecules A and B. The small dimer interface occurs through residues in α-helix F. (C) Details of the residues involved in the oligomeric interaction with Ser197, Val196, and Leu193 from molecule A interacting with His192, Leu193, and Val196 from molecule B. (D) Sedimentation velocity profiles of $Rep68_{wt}$, Rep68-L193AV196A, Rep68-H192A, and Rep68-S197A. Experiments were carried out at 20 °C and 25 000 rpm on a Beckman XL-I analytical ultracentrifuge. Scans were collected every 2 min using absorbance at 280 nm. Data was analyzed using the program Sedfit.[51] (E) Helicase assay of Rep68 mutants. Fluorescent-labeled DNA molecule has a 3′ single-stranded tail and an 18 bp region (first lane). Upon ATP addition, helicase activity is shown as the 18 nucleotide fluorescein-labeled ssDNA is displaced (second lane). Rep68, L193AV196A, H192A, and S197 proteins all unwind DNA. K340H mutant is ATPase-negative and does not unwind DNA. Last lane shows the DNA substrate after heating at 100 °C for 5 min. (F) Comparison of the production of rAAV2-GFP infectious particles in the presence of wt or LV mutant. Various volumes of supernatant (in (μ), x axis) from 293T cells producing rAAV2-GFP were added to HeLa cells, and the percentage of GFP-positive infected cells was determined by FACS analysis, as described in Materials and Methods. Data is presented as average ± standard deviation from three independent experiments.

Brookhaven National Laboratories at beamline X9 and at the Advanced Light Source at the Lawrence Berkeley National Laboratory at the SYBILS beamline. Three different concentrations of each sample were prepared and measured (1, 2, and 3 mg/mL). All data were processed with the package ATSAS.[52] Buffer subtraction was carried out using beamline-specific software. Radii of gyration (R_g) were evaluated using the Guinier approximation, $sR_g < 1.3$. Distance distribution functions and maximum diameters D_{max} were calculated using the program GNOM.[53] SAXS molecular envelopes were calculated using the programs DAMMIN and GASBOR.[54,55] Conformational flexibility of the linker and C-terminal tail was analyzed with the program EOM.[56] We used the structures of Rep40 (PDB ID: 1s9h) and of OBDL and connect the two domains with a flexible linker (residues 215−224) to generate an atomic model of AAV2 Rep68. Using this initial model, we performed rigid-body and molecular dynamics using BILBOMD to generate the best Rep68 model that fits the SAXS data.[57]

DOI: 10.1021/acs.biochem.5b00610
Biochemistry 2015, 54, 5907−5919

DNA Helicase Assay. The helicase assay was based on a modification of the strand-displacement assay described elsewhere.[58] The DNA substrate (28:18) consists of 3′ tail of 10 nucleotides adjacent to 18 bp. The top strand has been labeled with fluorescein (F) at the 5′ end. All reactions were performed in a final volume of 50 μL in a buffer containing 25 mM HEPES, 50 mM NaCl, pH. 7.0. For the reaction, 1 μM protein was mixed with 0.5 μM ds F-DNA (28:18). The reaction was started by addition of 5 mM ATP-$MgCl_2$ and 2.5 μM trap ssDNA. Reaction was incubated at 25 °C for 1 h. EDTA was used to stop the reaction at a final concentration of 20 mM. Aliquots of 10 μL were loaded in a 12% bis-acrylamide gel (30%) (19:1) using 6× loading dye (0.25 xylene cyanol FF, 30% glycerol). For densitometry and analysis of the bands, a Gel Doc EZ Imager was used, using the automatic lane and band detection tool. Lane background subtraction, white illumination, and an activation time of 300 s were used for the analysis.

Fluorescent Anisotropy Binding Assays. Binding assays were performed using 5 nM fluorescein labeled 41-mer DNA containing the Rep binding site (TGGCGGCGGTTGGG-GCTCGGC*GCTCGCTCGCTC*GCTGGGCG). Rep68 at different concentrations was mixed with DNA at a final volume of 300 μL using the following buffer: 25 mM HEPES (pH 7.0), 200 mM NaCl, 1 mM TCEP. Fluorescence readings were taken on a PC1 fluorimeter (ISS, Inc.) with excitation and emission filters at 492 and 528 respectively. Tubes were equilibrated at 20 °C for 20 min before measurement. Each anisotropy point is the average of 10 measurements. Anisotropy is calculated as the ratio of the difference between vertical and horizontal emission intensities over the total normalized intensity. The fraction of DNA bound (B) was calculated using eq 1

$$B = ([A]_x - [A]_{DNA})/([A]_{FINAL} - [A]_{DNA}) \quad (1)$$

where $[A]_x$ represents the anisotropy measured at protein concentration x, $[A]_{DNA}$ is the anisotropy of free fluorescence DNA, and $[A]_{FINAL}$ is the anisotropy at saturation. Data was fit to a single binding site model using the program PRISM6 (GraphPad). Each experiment was done in triplicate.

AAV Infectious Particles Assay. 293T cells were transfected with three plasmids using polyethylenimine (PEI): an AAV2 ITR-containing plasmid encoding a CAG-controlled GFP gene (pTRUF11), a helper plasmid expressing AAV2 Rep (wt or L193A-V196A cloned from the pHisRep68LV/15b) and Cap derived from pDG, and a third construct containing the adenovirus helper functions (HGTI plasmid).[59,60] The presence of the L193A−V196A double mutation was confirmed by sequencing (Eurofins). After 72 h, the cell supernatant was harvested and increasing volumes of supernatant were used to infect HeLa cells. The percentage of GFP-positive HeLa cells was determined 48 h postinfection by FACS (FACSCanto, BD Biosciences).

RESULTS

We opted to structurally characterize full-length AAV2 Rep68 using small-angle X-ray scattering and rigid-body modeling. A diagram of the different domains and protein constructs used in this study is shown in Figure 1A. We first determine the crystal structure of its OBD including the region spanning the linker, followed by SAXS studies of Rep40 (SF3 helicase domain) to study the dynamic behavior of the C-terminal tail. Finally, to determine the relative orientation of its two functional domains, we determined the SAXS solution structure of Rep68.

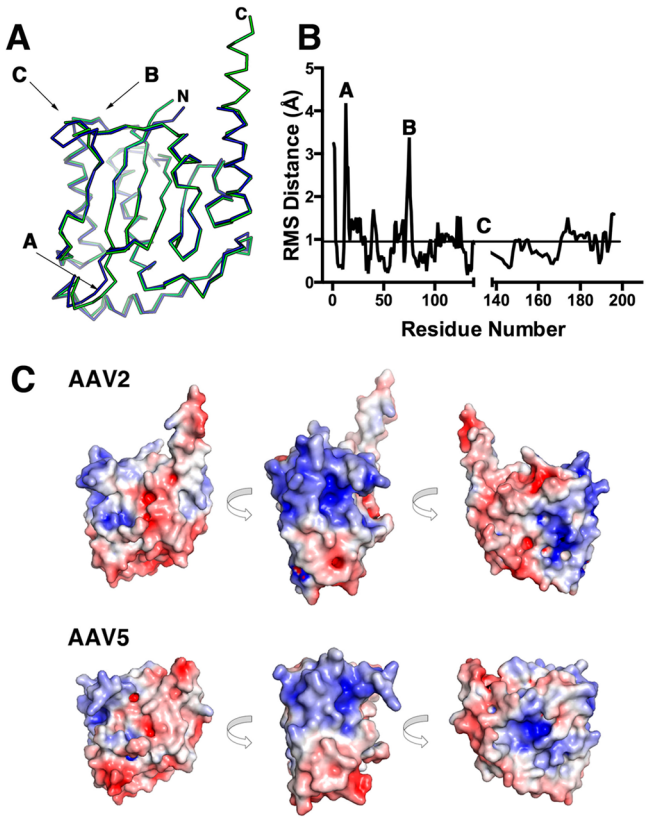


Figure 3. Comparison of AAV2 and AAV5 OBD structures. (A) Ribbon diagram of the superposition of the AAV2 (green) and AAV5 (blue) OBD structures. The N and C termini are marked. Regions that have the largest rmsd are indicated by arrows. (B) Plot of rmsd differences against residue number between AAV2 and AAV5 OBD. (C) Solvent-accessible electrostatic surface comparison of AAV2 and AAV5 OBDs showing three views of 90° counterclockwise rotations. Positive regions are shown in blue, and negative, in red.

Structure of AAV-2 Rep OBD. The OBD construct (residues 1−208) was initially obtained from limited proteolysis experiments.[61] The structure was solved by molecular replacement and refined to 2.3 Å resolution (Table 1). There are three molecules in the asymmetric unit: molecule A, spanning residues 1−138 and 143−206, molecule B, spanning residues 1−139 and 143−201, and molecule C, covering residues 1−131 and 141−191. The missing residues are all part of the DNA binding loop (L_{DB}). Molecules A and B form a dimer interacting through their C-terminal helix F, whereas molecule C does not make any close contacts. The OBD structure is characterized by a central five-stranded antiparallel β-sheet flanked by six α-helices, three at each side of the sheet (Figure 1B). The topology of the β-sheet follows a 4−1−3−2−5 order, with helices E and F on one side of the sheet and helices B−D positioned on the opposite side (Figure 1C,D). There are two long loops: the DNA binding loop L_{DB} connects β-strand 4 to α-helix E and loop L_{BC} connects α-helices B and C and protrudes underneath α-helix F (Figure 1C). L_{DB} was identified as being important in making specific contacts with the major groove of GCTC repeats and shows the highest degree of variation among all AAV Rep isoforms.[31,62] This is a dynamic region, as shown by the lack of electron density in all three molecules found in the asymmetric unit. The C-terminal helix αF is long, spanning residues 182−206, and protrudes at an approximately 45° angle from the plane of the β-sheet

DOI: 10.1021/acs.biochem.5b00610
Biochemistry 2015, 54, 5907−5919

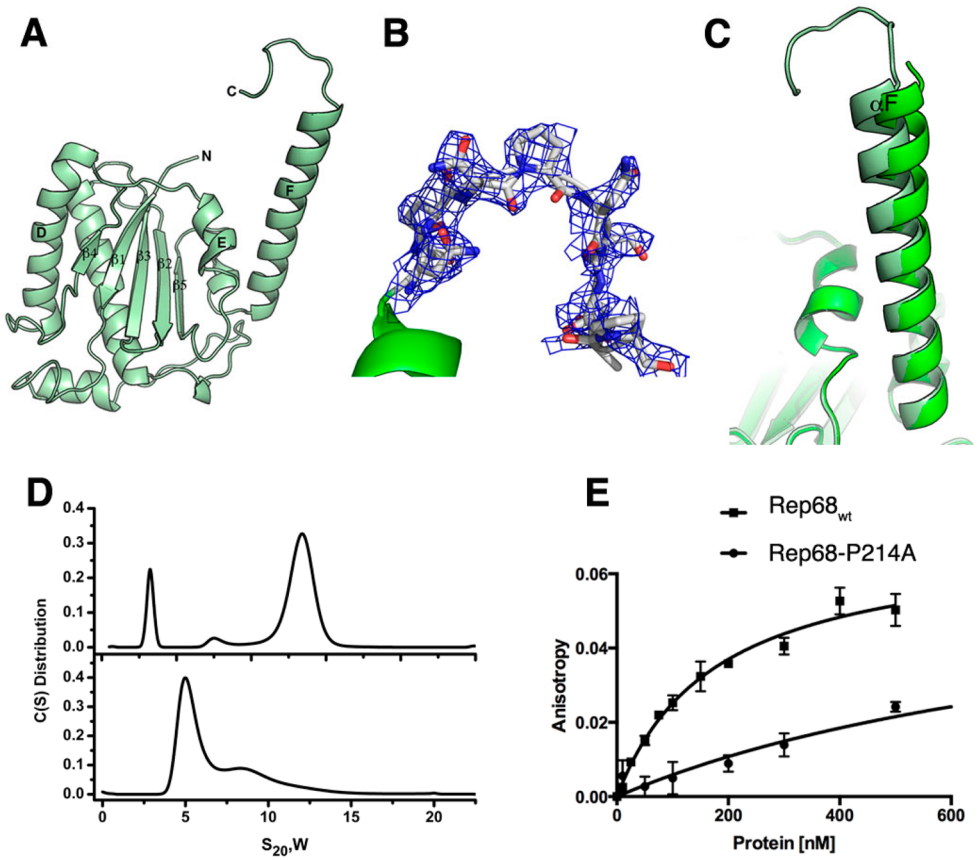


Figure 4. Structure of OBDL. (A) Ribbon diagram of the structure spanning residues 1−214. (B) Electron density of region 206−214 contoured at 0.9σ. The density fits most of the residues, but it suggests that the region is dynamic. (C) Superposition of OBD and OBDL structures illustrating the flexibility of helix F. (D) Comparison of the sedimentation velocity profiles between $Rep68_{wt}$ (top) and Rep68-P214A (bottom). (E) Comparison of binding isotherms of $Rep68_{wt}$ and Rep68-P214A. Fluorescent anisotropy binding assays were carried out with a 41-mer AAVS1 DNA labeled with fluorescein. Apparent dissociation constants were determined from triplicate experiments and fitted to a one binding site-specific binding model.

(Figure 1D). The residues at the end of helix F appear to be dynamic, as we could see only up to residue 201 in molecule B and to residue 191 in molecule C. Helix F is loosely packed against helix E and the loop connecting β5 to αF. The interactions that keep these three elements together are sparse, suggesting that helix F is flexible. Several dynamic regions in the AAV2 OBD can be observed by the superposition of the three molecules in the asymmetric unit (Figure 2A). The three molecules superimpose with an overall rmsd over 197 C_α of 0.62 Å. Two regions can be identified with larger than average values: region 1 includes resides 15−34 spanning helices A and B, and region 2 includes residues from L_{DB}. Not surprisingly, based on the AAV5 OBD−RBS structure, both of these regions are involved in DNA interactions: region 1 is involved in recognition of the ITR hairpin stem 2, and L_{DB} recognizes part of the GCTC repeat.[31] Thus, folding and stabilization of these elements must occur upon DNA binding. Two of the OBD molecules in the asymmetric unit form a dimer through interactions with residues present in α-helix F, resembling a pseudocoiled coil (Figure 2B). These include L193, V196, H192, and S197 making hydrophobic and hydrogen-bond interactions (Figure 2C). This type of interface was also observed in the AAV5 OBD−RBS complex, which may suggest a functional role for this interface in the context of full-length Rep78/Rep68 proteins.[31] To determine if these interactions are functionally relevant in Rep68 function and their effect on the AAV life cycle, we examined the ability of several Rep68 mutants to oligomerize, to unwind dsDNA, and to support the production of infectious particles. Figure 2D shows that Rep68 mutants H192A, S197A, and L193AV196A have similar sedimentation velocity profiles as that of $Rep68_{wt}$, forming the 13S peak that corresponds to heptameric/octameric ring species.[45] In addition, the mutants have similar unwinding activity as that of the wild-type protein (Figure 2E). Thus, the different mutants are functionally equivalent to $Rep68_{wt}$. This conclusion is supported by the ability of the Rep68-L193AV196A double mutant to produce infectious viral particles as efficiently as wild-type Rep68 (Figure 2F). These results show that the dimer interactions are generated during crystal packing and are not involved in Rep68 oligomerization and function.

Structural comparison of the AAV2 and AAV5 OBD. The Rep proteins from the majority of AAV serotypes show an overall homology with the AAV2 OBD of 95% (Figure S1). In contrast, AAV5 Rep has only 59% identity over the N-terminal 224 residues. However, the AAV2 and AAV5 OBD structures superimposed well, with an overall rmsd of 0.94 Å for 191 aligned C_α carbons. As shown in Figure 3A,B, there are, nonetheless, three regions that show differences: region A (residues 12−15) includes the linker connecting β1 to αA, region B (74−77) spans the turn connecting β3 to αD, and region C (139−147) consists of L_{DB}. Among all AAV serotypes,

DOI: 10.1021/acs.biochem.5b00610
Biochemistry 2015, 54, 5907−5919

AAV5 and AAV8 have the largest sequence variations. For instance, AAV5 has a shorter L_{DB} loop than AAV2, whereas in AAV8, the loop is longer by 2 residues. In addition, the AAV5 protein has four additional residues after position 198. Moreover, sequence conservation in the αF region is low in AAV5, suggesting that this structure may indeed be shorter than in AAV2 (Figure S1). The overall electrostatic surface representations of the two proteins look similar, but subtle differences are visible. In both structures, one side of the surface is highly positively charged; however, in AAV2 OBD, the positive patch is wider and more intense. This is seen more explicitly if we compare the positive surface potential values between the two OBDs, with 2130.6 kcal/molq for AAV2 and 1720.4 kcal/molq for AAV5.[63] This area includes L_{DB}, helix D, and the N-terminal half of α-helix C. The former two interact with the GCTC repeats, whereas the latter is involved in the recognition of the RBE′ hairpin in the ITRs.[31] Helix F is mostly negatively charged, but in AAV2, there is a slightly positive patch located in the first three turns that is not visible in AAV5 (Figure 3C). Because the structure of the AAV5 construct includes up to residue 197 and the sequence homology around this region is less than 80%, it is not apparent whether helix F in AAV5 will be as extended as in AAV2. Whether these structural and sequence variations found in L_{DB} and the linker region account for the differences in nicking specificity observed for AAV2 and AAV5 Rep proteins remains to be determined and will require structures with DNA substrates along the nicking reaction pathway.[64,65]

Crystal Structure of OBD-Linker. We and others have described previously that the linker region connecting the OBD and the SF3 helicase domain of the large Rep proteins is pivotal for the assembly of multimeric Rep protein complexes.[46,47] To better characterize this region, we expressed, purified, and crystallized a construct containing the OBD and the linker that we termed OBDL, spanning residues 1−224. OBDL crystals diffracted to 2.6 Å and belong to space group $P2_1$, with 11 molecules in the asymmetric unit. The electron density corresponding to the C-terminus of helix F and the linker region varies among the different molecules in the asymmetric unit; nevertheless, we were able to build a model of OBDL up to residue 214 (Figure 4). We did not detect any electron density from residues 215−224. The structure shows that helix F spans to residue 205 and is followed by a loop that makes a downward turn toward the core of the protein at proline 209 (Figure 4B). Superposition of OBD with OBDL shows that helix F is flexible and moves as a rigid body with respect to the core of the protein, pivoting around the loop that connects it with strand β5. While the two OBD structures superimpose with an overall rmsd of ∼0.8 Å, the two αF helices superimpose with an rmsd of only 1.3 Å and appear at a different angle with respect to the core of the protein, with a difference between them of ∼2.4° (Figure 4C). The OBDL structure also implies that the linker is not in an extended conformation but, instead, has a particular structure that is constrained by the two prolines in the linker. This particular conformation may bring the two domains together in the context of full-length Rep68. To test the importance of the proline residues, we mutated proline 214 to alanine and measured the effect of this on Rep68's properties. Figure 4D shows that this mutation drastically alters the oligomeric behavior of Rep68 and affects its ability to form the 13S oligomer species (Figure 4D). Moreover, the mutation affects the ability of Rep68 to bind double-stranded DNA containing a Rep binding site (RBS). Figure 4E shows that the Rep68-P214A mutant binds DNA ∼ 7 times less than wild-type protein. Taken together, our results show that the conformation of the linker is critical for the function of Rep68.

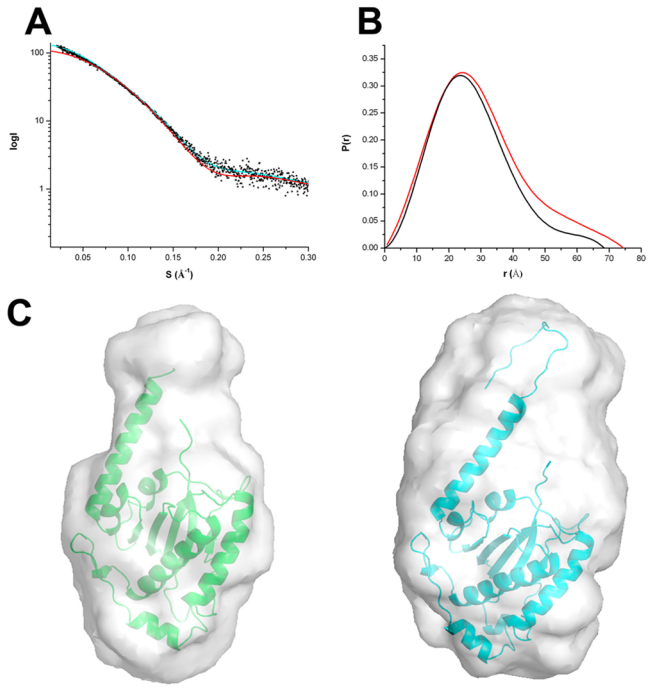


Figure 5. SAXS analysis for OBD and OBDL. (A) SAXS scattering data from OBD (blue) and OBDL (red) in solution. Superimposed is the scattering calculated from the atomic models for OBD (■) and OBDL (△). (B) Normalized pair distribution functions $P(r)$ for OBD (black) and OBDL (red) calculated from the experimental scattering curves using the program PRIMUS. (C) Low-resolution envelopes of DAMMIN and GASBOR models for OBD (left) and OBDL (right) are superimposed with the X-ray structure of OBD and OBDL, respectively.

SAXS Studies of OBDL Validate the Conformation of the Linker. To gain further structural information about the linker domain and to validate the OBDL X-ray structure, we performed solution studies on OBD and OBDL. The concentration of NaCl in the buffer was kept at 0.5−1 M to prevent formation of oligomers induced by the presence of the linker.[46] The calculated sedimentation coefficient for both constructs is ∼2.0 S (Figure S2). The small difference in S value suggests that the linker region is not in an extended conformation. However, the smaller sedimentation velocity coefficient of OBDL implies a slightly more elongated shape. We subjected both constructs to SAXS studies under conditions that generated single species, as determined by sedimentation velocity. Scattering profiles from three different concentrations were collected, and a final scattering curve was obtained by merging the best curves using the program Almerge.[66] The scattering profiles for OBD and OBDL are shown in Figure 5A. The merged data was used to calculate both the radius of gyration (R_g) and the $P(r)$ distribution function as described in the Material and Methods (Figure 5B). *Ab initio* models were produced using the programs DAMMIN and GASBOR. For each protein, 10 independent models were generated and averaged with Damaver. The final models from the two independent reconstructions were aligned using the program Supcomb, resulting in an overall normalized spatial discrepancy (NSD) value of 0.54, suggesting an excellent correlation between the two independent reconstructions.[67] The resulting GASBOR models for OBD and OBDL are shown

DOI: 10.1021/acs.biochem.5b00610
Biochemistry 2015, 54, 5907−5919

Table 2. Hydrodynamic Parameters

parameter	OBD (1−208)	OBDL (1−224)	Rep40$_{wt}$ (225−536)	Rep68Y224AΔ (1−490)
S (s)[a]	2.1	1.97	nd[d]	nd
R_g (nm)[b]	21.5 ± 0.3	23.2 ± 2.3	27.6 ± 3.8	38.9 ± 0.1
D_{max} (nm)[c]	68.2	74.4 (104.96)	98.8	123.8
χ	1.0[e]	1.9[e]	0.5[f]	0.8[g]

[a]Determined from sedimentation velocity data. [b]Determined from Guinier analysis. [c]Determined from $P(r)$ analysis. [d]nd, not determined. [e]Experimental/model data fit using FoXS.[74] [f]Experimental/model data fit using EOM. [g]Experimental/model data fit using CORAL.

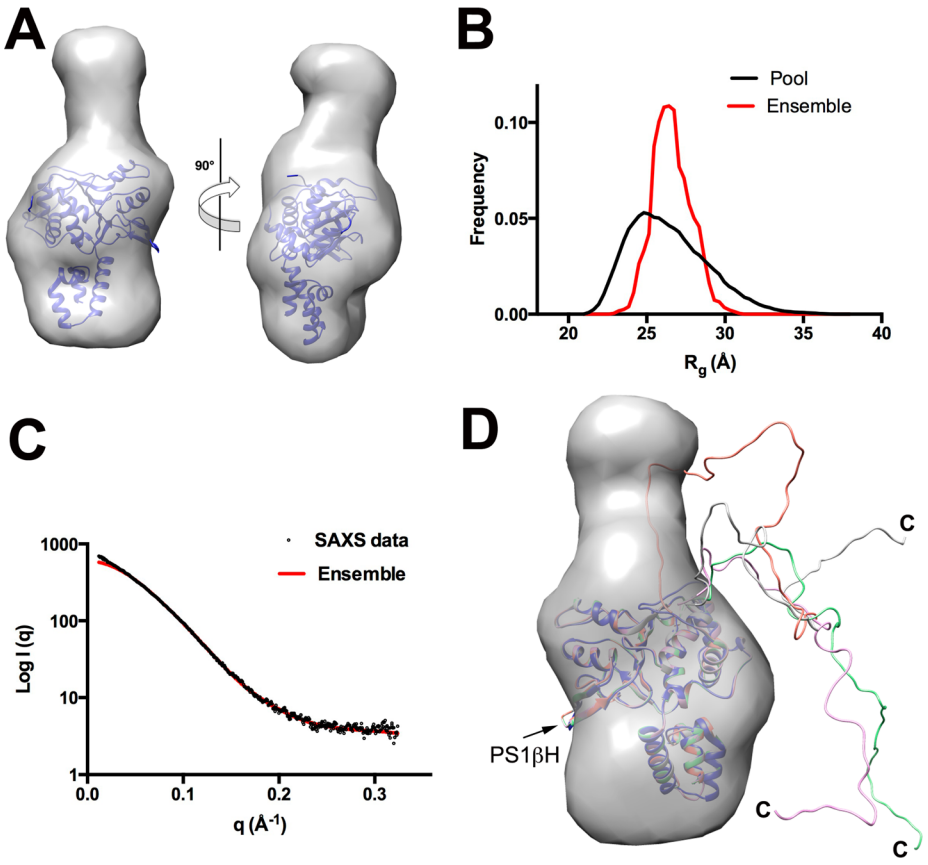


Figure 6. SAXS analysis of Rep40/Rep68 C-terminal tail. (A) Shape reconstruction of Rep40$_{wt}$ and the docked X-ray structure of Rep40. (B) Comparison of the R_g distribution of the initial EOM generated pool (black line) and the selected model ensemble (red line). (C) Scattering curve of Rep40$_{wt}$ (black) and fit of the optimal EOM model ensemble (red). (D) Superposition of the ensemble generated model showing the preferred conformations of the C-terminal tails.

in Figure 5C with the superimposed X-ray structures of the corresponding constructs. The calculated R_g and D_{max} values are similar for both models (Table 2). The OBDL GASBOR *ab initio* model shows an envelope that resembles a round cylinder, suggesting that, under our experimental conditions, the region between 208 and 224 is not extended but may resemble the overall conformation seen in the OBDL crystal structure (Figure 5C, right panel). To further corroborate the hypothesis that the linker is neither extended nor highly flexible, we carried out molecular dynamics using BILBOMD to determine if an ensemble with multiple conformations of the linker described the SAXS scattering profile of OBDL.[57] The results show that inclusion of more than one conformation does not improve the fit to the experimental data significantly (data not shown).

Modeling Flexibility of C-Terminal Tail of Rep68. AAV Rep proteins Rep40 and Rep68 have a C-terminal tail of 46 residues that extends from the core of the helicase domain, as determined by limited proteolysis experiments.[38] To determine the overall flexibility of the C-terminal tail, we carried out SAXS experiments of Rep40wt at different concentrations, and data were analyzed as described previously. A GASBOR model of Rep40 clearly resembles the overall shape of the helicase domain and has a characteristic flat disk shape with an additional elongated density that is not represented in the X-ray structure, which should correspond to the C-terminal tail (Figure 6A). However, the dimensions are smaller than those expected from a fully extended C-terminal tail. To analyze the flexibility of the tail, we used EOM software to search for an optimal ensemble of conformations that will best fit the scattering data. This method generates a large pool of models (10 000) with random conformations and uses a genetic algorithm to select an optimized ensemble of configurations that best describes the SAXS scattering data (Figure 6B).[56,68] The best ensemble consisted of four models that fit the experimental curve with a χ of 0.5 (Figure 6C). However, the

DOI: 10.1021/acs.biochem.5b00610
Biochemistry 2015, 54, 5907−5919

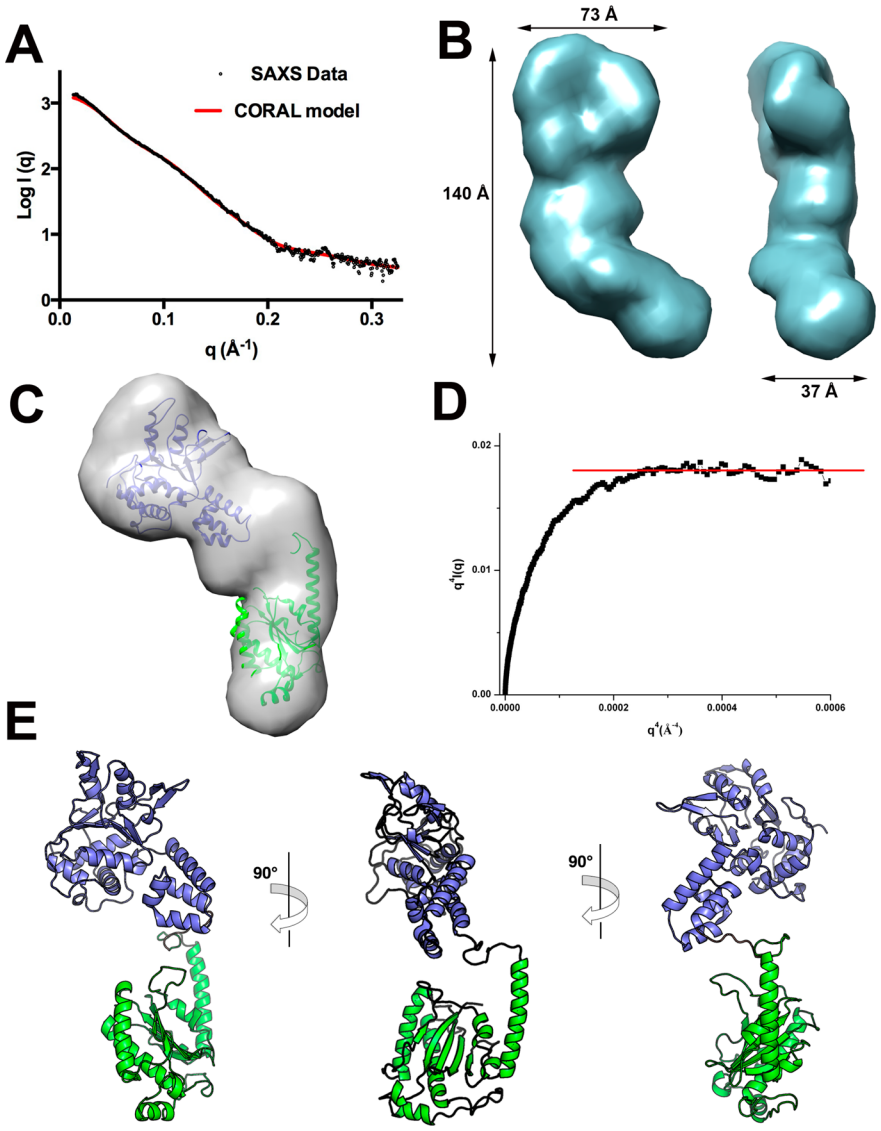


Figure 7. SAXS modeling of Rep68. (A) Fit of the experimental SAXS data (black circles) with the theoretical scattering profile obtained from CORAL model (red line) with a χ value of 0.8. (B) Two views of the GASBOR averaged molecular envelope for Rep68Δ showing the approximate dimensions in angstroms. (C) Docking of OBDL and Rep40 atomic structures to the GASBOR *ab initio* envelope. (D) Porod–Debye plot of Rep68Y224AΔ SAXS data (black squares) supporting a compact protein with little flexibility. Red line represents the linear plateau. (E) Three views of the CORAL rigid-body final model of Rep68Y224AΔ.

narrowness of the ensemble R_g distribution suggests that a limited number of conformations are preferred, with one configuration in particular accounting for about 50% of the population. Interestingly, all of the selected models have the C-terminal oriented toward the opposite site of the presensor 1 β-hairpin (PS1βH), a motif that is involved in DNA interactions during DNA translocation and unwinding (Figure 6D).[30,69] Thus, our study shows that the C-terminal tail, although flexible, has a preferred set of dominant conformations.

SAXS-Based Structural Modeling of Rep68 Shows a Compact Conformation in Solution. We previously showed that Rep68 has complex dynamic oligomeric behavior in solution and is present as a mixture of multiple oligomers, including heptameric and octameric rings.[45] To obtain a homogeneous population of Rep68 monomers, we took advantage of the Y224A mutation that affects its tendency to form multiple oligomeric species.[46] The Rep68Y224A protein is present as a monomer at concentrations less than 5 mg/mL and under high salt conditions. Moreover, we have determined that this mutation has helicase activity comparable to that of the wild-type protein and does not cause any significant structural change.[70] To facilitate model building and data interpretation, we used a truncated Rep68 (1–490) to eliminate the C-terminal tail residues.[38] SAXS data was collected at 2, 3, and 4 mg/mL and processed as described in the previous sections (Figure 7A). The GASBOR generated model produces an elongated prolate ellipsoid that is slightly curved with dimensions 140 × 73 × 37 Å^3 (Figure 7B). Individual atomic structures of OBDL and Rep40 can be easily docked into the envelope; in particular, one end of the particle has the shape of a flat disk that resembles the GASBOR Rep40 *ab initio* model (Figure 7C). We performed rigid-body modeling using the X-ray structures of OBDL (1–214) and Rep40 (225–490) with CORAL.[71] The domain boundaries define an unstructured linker spanning residues 215–224. However, secondary structure prediction suggests that the region between 220 and

DOI: 10.1021/acs.biochem.5b00610
Biochemistry 2015, 54, 5907–5919

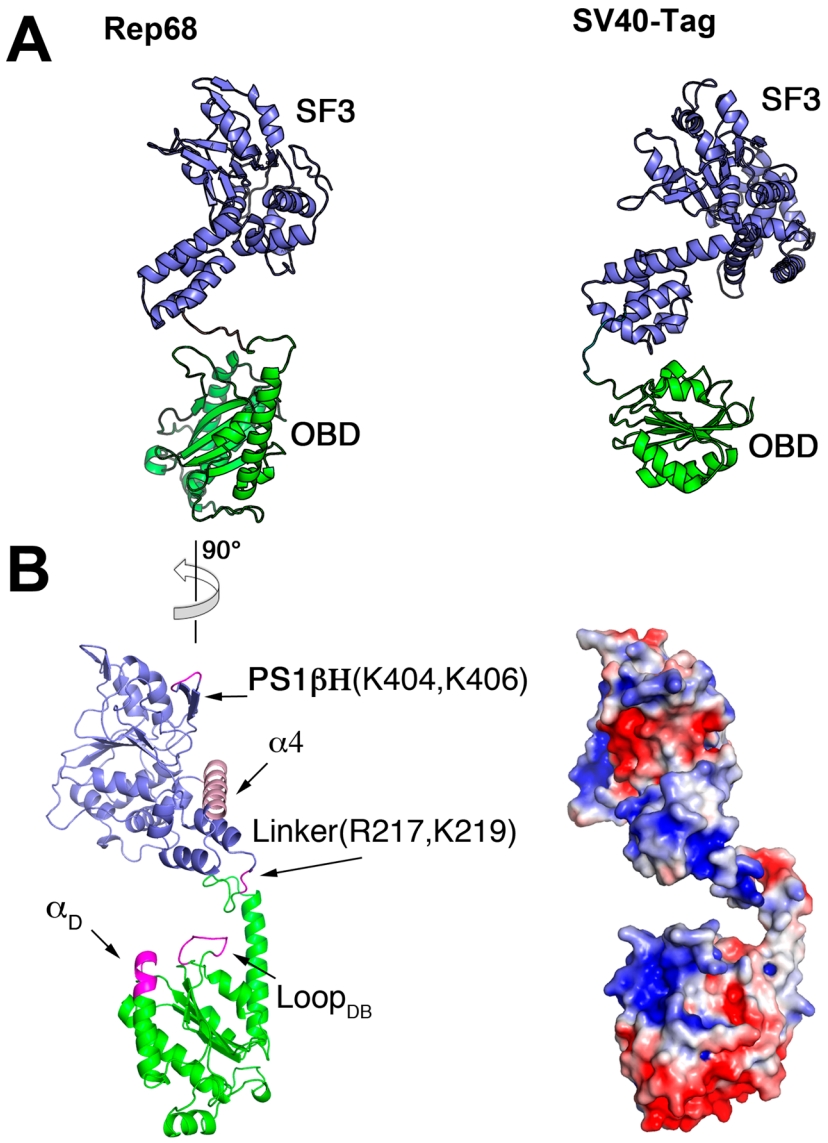


Figure 8. Rep68 DNA interacting face. (A) Comparison of Rep68 model (left) and SV40-LTag (right). Colored equivalently are the OBD domain (green) and the helicase domain (blue). (B) (left) Rep68 structure showing the regions that interact with DNA colored in pink. Orientation of Rep68 is 90° counterclockwise from that in panel A. (right) Surface representation of Rep68 showing electropositive regions in blue and negative regions in red. View is in the same orientation as that on the left.

224 could form a α-helix extending into the first helix of the helicase domain.[46] Consequently, we generated multiple models that extended this helix at different positions. The Rep68 models fit the experimental data very well, with χ values in the range of 1.1−0.8. All models show that the two domains are positioned such that the long axis of the helicase domain is almost perpendicular to the OBD, as shown in Figure 7E. The best model was obtained by extending the N-terminal helicase helix to residue 223. Two main observations can be drawn from this model. First, the orientation of the two domains results in a Rep68 structure where the motifs that interact with DNA in both the OBD (L_{DB} and helix C) and helicase domain (β-hairpin 1) are on the same face of the protein. Second, the two domains are closely positioned, making an extended linker structure unlikely. To further support this conclusion, we assessed interdomain flexibility of Rep68 using two methodologies. First, a Porod−Debye plot of the scattering data shows a plateau that is a signature for a compact molecule.[72] The data fits the linear region with a Porod coefficient of 4, again consistent with a compact molecule (Figure 7D). In addition, we carried out BILBOMD and determined that a single model fits the data equally as well as with multiple conformations (Figure 7A).

DISCUSSION

Our results show new structural features that increase our knowledge of the architecture of AAV Rep proteins. The combined X-ray and SAXS studies show that the AAV2 Rep68 linker region is partially structured, with helix F of the OBD extending until residue 209 and protruding from the main core at a 45° angle. The presence of two proline residues (209 and 214) seems to impart a certain rigidity to the loop region, making a small turn before continuing toward the helicase domain (Figure 4). This conformation is important for Rep68 function, as mutation of one of the proline residues in the linker (P214) is sufficient to prevent proper oligomerization and DNA binding. In addition, part of the remaining linker may extend the first helix of the helical domain of Rep52/40.[46] Both

DOI: 10.1021/acs.biochem.5b00610
Biochemistry 2015, 54, 5907−5919

of these structural features bring the OBD and helicase domain closer to each other. However, although the linker is not extended, it allows for a certain degree of flexibility. Our rigid-body modeling generated different models with slightly different conformations of the two domains that fit the scattering data equally well (Figure S3). This suggests a certain degree of conformational flexibility between the two domains. Structurally, this flexibility originates from the linker, the OBD helix F, and the helical bundle of the helicase domain. Alignment of the OBD and OBDL structures shows that helix F can pivot and move relative to the main core of the domain (Figure 4C). Likewise, alignment of the three Rep40 molecules found in the asymmetric unit of the X-ray structure (PDB ID: 1s9h) shows the helical bundle at different positions relative to the AAA^+ domain (data not shown). This flexibility is important to accommodate changes occurring upon DNA binding and oligomerization. Moreover, analysis of the flexibility of the C-terminal tail in Rep68 suggest that there is a preference in the conformations acquired by the C-terminal tail that positions it at the opposite end of the β-hairpin. This is important in the context of formation of Rep68 oligomeric rings because other conformations may produce steric clashes and inhibit their formation.

Finally, our Rep68 model closely resembles the configuration of the helicase and OBD in one of the molecules seen in the X-ray structure of the SV40 large T antigen (SV40-LTag) in complex with DNA.[73] This arrangement results in critical residues that interact with DNA being located on one face of the protein (Figure 8B). In our Rep68 model, these regions include loop L_{DB} and helix α_D in the OBD, the helicase domain (PS1β residues K404 and K406), and linker residues R217 and K219 that have been shown to play a role in complex formation and DNA binding in AAV5.[47] In addition, the SV40-LTag structure shows that the helical bundle (Zn domain in SV40-LTag) interacts with DNA, docking into the major groove.[73] Our model shows that the position of the helical bundle is similar, and it is intriguing to postulate that the conformation of Rep68 is prealigned to interact with DNA in way similar to that of SV40-LTag (Figure 8A).

In conclusion, we have completed the first structural description of full-length AAV2 Rep68 protein, revealing the orientation and relative position of its functional domains. These observations provide new clues to explain its DNA binding mode; however, our model does not discard possible changes that should occur upon DNA binding and/or oligomerization. These questions require future structural studies of high-resolution structures of Rep68 oligomers alone and in complex with DNA.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.biochem.5b00610.

Sequence and structural alignment of the AAV OBD serotypes, sedimentation velocity profiles of AAV2 OBD and OBDL, and superposition of different Rep68 models obtained by independent CORAL runs (PDF).

■ AUTHOR INFORMATION

Corresponding Author

*Tel.: +1 (804) 628-1202; Fax: +1 (804) 628-3501; E-mail: Carlos.Escalante@vcuhealth.org.

Funding

This work was supported by NIH grant R01-GM092854 (C.R.E.) and UK MRC grant 1001764 (R.M.L.).

Author Contributions

[@]These authors contributed equally to this study.

Notes

The authors declare no competing financial interest.

Accession Codes

PDB ID Codes: 4ZO0, 5DCX

■ ACKNOWLEDGMENTS

We would like to thank members of NSLS beamlines X6a (Vivian Stojanoff, Jean Jakoncic, and Edwin Lazo) and X9 (Lin Yang and Marc Allaire) and ALS SIBYLS beamline (Kevin Dyer and Kathryn Burnett) for help in data collection. We thank Vishaka Santosh for helpful comments on the preparation of this manuscript.

■ ABBREVIATIONS

AAV, adeno-associated virus; OBD, origin binding domain; OBDL, origin binding domain with linker; HUH, histidine−hydrophobic−histidine; DBD, DNA binding domain; SF3, superfamily 3 helicase; SAXS, small-angle X-ray scattering; ORF, open reading frame; ITR, inverted terminal repeat; RBS, Rep binding site; ssDNA, single-stranded DNA; AAA^+, ATPases associated with various cellular activities; IPTG, isopropyl-β-D-thiogalactopyranoside; LB, Luria−Bertani; TCEP, tris(2-carboxyethyl)phosphine; NTA, nitrilotriacetic acid; MPD, 2-methyl-2,4-pentanediol; NSLS, National Synchrotron Light Source; R_g, radius of gyration, PEI, polyethylenimine; FACS, fluorescent activated cell sorting

■ REFERENCES

(1) Henckaerts, E., and Linden, R. M. (2010) Adeno-associated virus: a key to the human genome? *Future Virol. 5*, 555−574.

(2) Weitzman, M. D., and Linden, R. M. (2012) Adeno-associated virus biology. *Methods Mol. Biol. 807*, 1−23.

(3) Lusby, E., Fife, K. H., and Berns, K. I. (1980) Nucleotide sequence of the inverted terminal repetition in adeno-associated virus DNA. *J. Virol. 34*, 402−409.

(4) Samulski, R. J., Srivastava, A., Berns, K. I., and Muzyczka, N. (1983) Rescue of adeno-associated virus from recombinant plasmids: gene correction within the terminal repeats of AAV. *Cell 33*, 135−143.

(5) Senapathy, P., Tratschin, J. D., and Carter, B. J. (1984) Replication of adeno-associated virus DNA. Complementation of naturally occurring rep- mutants by a wild-type genome or an ori-mutant and correction of terminal palindrome deletions. *J. Mol. Biol. 179*, 1−20.

(6) Beaton, A., Palumbo, P., and Berns, K. I. (1989) Expression from the adeno-associated virus p5 and p19 promoters is negatively regulated in trans by the rep protein. *J. Virol. 63*, 4450−4454.

(7) Pereira, D. J., McCarty, D. M., and Muzyczka, N. (1997) The adeno-associated virus (AAV) Rep protein acts as both a repressor and an activator to regulate AAV transcription during a productive infection. *J. Virol. 71*, 1079−1088.

(8) Surosky, R. T., Urabe, M., Godwin, S. G., McQuiston, S. A., Kurtzman, G. J., Ozawa, K., and Natsoulis, G. (1997) Adeno-associated virus Rep proteins target DNA sequences to a unique locus in the human genome. *J. Virol. 71*, 7951−7959.

(9) Young, S. M., Jr., and Samulski, R. J. (2001) Adeno-associated virus (AAV) site-specific recombination does not require a Rep-dependent origin of replication within the AAV terminal repeat. *Proc. Natl. Acad. Sci. U. S. A. 98*, 13525−13530.

(10) McCarty, D. M., Ryan, J. H., Zolotukhin, S., Zhou, X., and Muzyczka, N. (1994) Interaction of the adeno-associated virus Rep

DOI: 10.1021/acs.biochem.5b00610
Biochemistry 2015, 54, 5907−5919

protein with a sequence within the A palindrome of the viral terminal repeat. *J. Virol. 68*, 4998−5006.
(11) Hauswirth, W. W., and Berns, K. I. (1977) Origin and termination of adeno-associated virus DNA replication. *Virology 78*, 488−499.
(12) Spear, I. S., Fife, K. H., Hauswirth, W. W., Jones, C. J., and Berns, K. I. (1977) Evidence for two nucleotide sequence orientations within the terminal repetition of adeno-associated virus DNA. *J. Virol. 24*, 627−634.
(13) Becerra, S. P., Rose, J. A., Hardy, M., Baroudy, B. M., and Anderson, C. W. (1985) Direct mapping of adeno-associated virus capsid proteins B and C: a possible ACG initiation codon. *Proc. Natl. Acad. Sci. U. S. A. 82*, 7919−7923.
(14) Becerra, S. P., Koczot, F., Fabisch, P., and Rose, J. A. (1988) Synthesis of adeno-associated virus structural proteins requires both alternative mRNA splicing and alternative initiations from a single transcript. *J. Virol. 62*, 2745−2754.
(15) Cassinotti, P., Weitzand, M., and Tratschin, J. D. (1988) Organization of the adeno-associated virus (AAV) capsid gene: mapping of a minor spliced mRNA coding for virus capsid protein. *Virology 167*, 176−184.
(16) Trempe, J. P., and Carter, B. J. (1988) Alternate mRNA splicing is required for synthesis of adeno-associated virus VP1 capsid protein. *J. Virol. 62*, 3356−3363.
(17) Muralidhar, S., Becerra, S. P., and Rose, J. A. (1994) Site-directed mutagenesis of adeno-associated virus type 2 structural protein initiation codons: effects on regulation of synthesis and biological activity. *J. Virol. 68*, 170−176.
(18) Sonntag, F., Schmidt, K., and Kleinschmidt, J. A. (2010) A viral assembly factor promotes AAV2 capsid formation in the nucleolus. *Proc. Natl. Acad. Sci. U. S. A. 107*, 10220−10225.
(19) Green, M. R., and Roeder, R. G. (1980) Transcripts of the adeno-associated virus genome: mapping of the major RNAs. *J. Virol. 36*, 79−92.
(20) Green, M. R., Straus, S. E., and Roeder, R. G. (1980) Transcripts of the adenovirus-associated virus genome: multiple polyadenylated RNAs including a potential primary transcript. *J. Virol. 35*, 560−565.
(21) Laughlin, C. A., Westphal, H., and Carter, B. J. (1979) Spliced adenovirus-associated virus RNA. *Proc. Natl. Acad. Sci. U. S. A. 76*, 5567−5571.
(22) Kyostio, S. R., Wonderling, R. S., and Owens, R. A. (1995) Negative regulation of the adeno-associated virus (AAV) P5 promoter involves both the P5 rep binding site and the consensus ATP-binding motif of the AAV Rep68 protein. *J. Virol. 69*, 6787−6796.
(23) Dutheil, N., Smith, S. C., Agundez, L., Vincent-Mistiaen, Z. I., Escalante, C. R., Linden, R. M., and Henckaerts, E. (2014) Adeno-associated virus Rep represses the human integration site promoter by two pathways that are similar to those required for the regulation of the viral p5 promoter. *J. Virol. 88*, 8227−8241.
(24) Im, D. S., and Muzyczka, N. (1990) The AAV origin binding protein Rep68 is an ATP-dependent site-specific endonuclease with DNA helicase activity. *Cell 61*, 447−457.
(25) Chiorini, J. A., Weitzman, M. D., Owens, R. A., Urcelay, E., Safer, B., and Kotin, R. M. (1994) Biologically active Rep proteins of adeno-associated virus type 2 produced as fusion proteins in Escherichia coli. *J. Virol. 68*, 797−804.
(26) Weitzman, M. D., Kyostio, S. R., Kotin, R. M., and Owens, R. A. (1994) Adeno-associated virus (AAV) Rep proteins mediate complex formation between AAV DNA and its integration site in human DNA. *Proc. Natl. Acad. Sci. U. S. A. 91*, 5808−5812.
(27) Urcelay, E., Ward, P., Wiener, S. M., Safer, B., and Kotin, R. M. (1995) Asymmetric replication in vitro from a human sequence element is dependent on adeno-associated virus Rep protein. *J. Virol. 69*, 2038−2046.
(28) Lamartina, S., Ciliberto, G., and Toniatti, C. (2000) Selective cleavage of AAVS1 substrates by the adeno-associated virus type 2 rep68 protein is dependent on topological and sequence constraints. *Journal of virology 74*, 8831−8842.
(29) King, J. A., Dubielzig, R., Grimm, D., and Kleinschmidt, J. A. (2001) DNA helicase-mediated packaging of adeno-associated virus type 2 genomes into preformed capsids. *EMBO J. 20*, 3282−3291.
(30) Yoon-Robarts, M., Blouin, A. G., Bleker, S., Kleinschmidt, J. A., Aggarwal, A. K., Escalante, C. R., and Linden, R. M. (2004) Residues within the B′ motif are critical for DNA binding by the superfamily 3 helicase Rep40 of adeno-associated virus type 2. *J. Biol. Chem. 279*, 50472−50481.
(31) Hickman, A. B., Ronning, D. R., Perez, Z. N., Kotin, R. M., and Dyda, F. (2004) The nuclease domain of adeno-associated virus rep coordinates replication initiation using two distinct DNA recognition interfaces. *Mol. Cell 13*, 403−414.
(32) Yoon-Robarts, M., and Linden, R. M. (2003) Identification of active site residues of the adeno-associated virus type 2 Rep endonuclease. *J. Biol. Chem. 278*, 4912−4918.
(33) Owens, R. A., Weitzman, M. D., Kyostio, S. R., and Carter, B. J. (1993) Identification of a DNA-binding domain in the amino terminus of adeno-associated virus Rep proteins. *J. Virol. 67*, 997−1005.
(34) Hickman, A. B., Ronning, D. R., Kotin, R. M., and Dyda, F. (2002) Structural unity among viral origin binding proteins: crystal structure of the nuclease domain of adeno-associated virus Rep. *Mol. Cell 10*, 327−337.
(35) Koonin, E. V., and Ilyina, T. V. (1993) Computer-assisted dissection of rolling circle DNA replication. *BioSystems 30*, 241−268.
(36) Chandler, M., de la Cruz, F., Dyda, F., Hickman, A. B., Moncalian, G., and Ton-Hoang, B. (2013) Breaking and joining single-stranded DNA: the HUH endonuclease superfamily. *Nat. Rev. Microbiol. 11*, 525−538.
(37) Gonzalez-Prieto, C., Agundez, L., Linden, R. M., and Llosa, M. (2013) HUH site-specific recombinases for targeted modification of the human genome. *Trends Biotechnol. 31*, 305−312.
(38) James, J. A., Escalante, C. R., Yoon-Robarts, M., Edwards, T. A., Linden, R. M., and Aggarwal, A. K. (2003) Crystal structure of the SF3 helicase from adeno-associated virus type 2. *Structure 11*, 1025−1035.
(39) Gorbalenya, A. E., Koonin, E. V., and Wolf, Y. I. (1990) A new superfamily of putative NTP-binding domains encoded by genomes of small DNA and RNA viruses. *FEBS Lett. 262*, 145−148.
(40) Yang, Q., Kadam, A., and Trempe, J. P. (1992) Mutational analysis of the adeno-associated virus rep gene. *J. Virol. 66*, 6058−6069.
(41) Walker, S. L., Wonderling, R. S., and Owens, R. A. (1997) Mutational analysis of the adeno-associated virus type 2 Rep68 protein helicase motifs. *J. Virol. 71*, 6996−7004.
(42) Wonderling, R. S., Kyostio, S. R., and Owens, R. A. (1995) A maltose-binding protein/adeno-associated virus Rep68 fusion protein has DNA-RNA helicase and ATPase activities. *J. Virol. 69*, 3542−3548.
(43) Smith, R. H., and Kotin, R. M. (1998) The Rep52 gene product of adeno-associated virus is a DNA helicase with 3′-to-5′ polarity. *J. Virol. 72*, 4874−4881.
(44) Mansilla-Soto, J., Yoon-Robarts, M., Rice, W. J., Arya, S., Escalante, C. R., and Linden, R. M. (2009) DNA structure modulates the oligomerization properties of the AAV initiator protein Rep68. *PLoS Pathog. 5*, e1000513.
(45) Zarate-Perez, F., Mansilla-Soto, J., Bardelli, M., Burgner, J. W., 2nd, Villamil-Jarauta, M., Kekilli, D., Samso, M., Linden, R. M., and Escalante, C. R. (2013) Oligomeric properties of adeno-associated virus Rep68 reflect its multifunctionality. *Journal of virology 87*, 1232−1241.
(46) Zarate-Perez, F., Bardelli, M., Burgner, J. W., 2nd, Villamil-Jarauta, M., Das, K., Kekilli, D., Mansilla-Soto, J., Linden, R. M., and Escalante, C. R. (2012) The interdomain linker of AAV-2 Rep68 is an integral part of its oligomerization domain: role of a conserved SF3 helicase residue in oligomerization. *PLoS Pathog. 8*, e1002764.
(47) Maggin, J. E., James, J. A., Chappie, J. S., Dyda, F., and Hickman, A. B. (2012) The amino acid linker between the endonuclease and helicase domains of adeno-associated virus type 5 Rep plays a critical role in DNA-dependent oligomerization. *Journal of virology 86*, 3337−3346.

DOI: 10.1021/acs.biochem.5b00610
Biochemistry 2015, 54, 5907−5919

(48) Otwinowski, Z., and Minor, W. (1997) Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol. 276*, 307−326.
(49) Adams, P. D., Grosse-Kunstleve, R. W., Hung, L. W., Ioerger, T. R., McCoy, A. J., Moriarty, N. W., Read, R. J., Sacchettini, J. C., Sauter, N. K., and Terwilliger, T. C. (2002) PHENIX: building new software for automated crystallographic structure determination. *Acta Crystallogr., Sect. D: Biol. Crystallogr. 58*, 1948−1954.
(50) Emsley, P., and Cowtan, K. (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr., Sect. D: Biol. Crystallogr. 60*, 2126−2132.
(51) Schuck, P. (2000) Size-distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and lamm equation modeling. *Biophys. J. 78*, 1606−1619.
(52) Konarev, P. V., Petoukhov, M. V., Volkov, V. V., and Svergun, D. I. (2006) ATSAS 2.1, a program package for small-angle scattering data analysis. *J. Appl. Crystallogr. 39*, 277−286.
(53) Svergun, D. (1992) Determination of the regularization parameter in indirect-transform methods using perceptual criteria. *J. Appl. Crystallogr. 25*, 495−503.
(54) Svergun, D. I. (1999) Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing. *Biophys. J. 76*, 2879−2886.
(55) Svergun, D. I., Petoukhov, M. V., and Koch, M. H. (2001) Determination of domain structure of proteins from X-ray solution scattering. *Biophys. J. 80*, 2946−2953.
(56) Bernado, P., Mylonas, E., Petoukhov, M. V., Blackledge, M., and Svergun, D. I. (2007) Structural characterization of flexible proteins using small-angle X-ray scattering. *J. Am. Chem. Soc. 129*, 5656−5664.
(57) Pelikan, M., Hura, G. L., and Hammel, M. (2009) Structure and flexibility within proteins as identified through small angle X-ray scattering. *Gen. Physiol. Biophys. 28*, 174−189.
(58) Amaratunga, M., and Lohman, T. M. (1993) Escherichia coli rep helicase unwinds DNA by an active mechanism. *Biochemistry 32*, 6815−6820.
(59) Zolotukhin, S., Potter, M., Hauswirth, W. W., Guy, J., and Muzyczka, N. (1996) A "humanized" green fluorescent protein cDNA adapted for high-level expression in mammalian cells. *J. Virol. 70*, 4646−4654.
(60) Grimm, D., Kern, A., Rittner, K., and Kleinschmidt, J. A. (1998) Novel tools for production and purification of recombinant adenoassociated virus vectors. *Hum. Gene Ther. 9*, 2745−2760.
(61) Yoon, M., Smith, D. H., Ward, P., Medrano, F. J., Aggarwal, A. K., and Linden, R. M. (2001) Amino-terminal domain exchange redirects origin-specific interactions of adeno-associated virus rep78 in vitro. *Journal of virology 75*, 3230−3239.
(62) Chiorini, J. A., Afione, S., and Kotin, R. M. (1999) Adeno-associated virus (AAV) type 5 Rep protein cleaves a unique terminal resolution site compared with other AAV serotypes. *J. Virol. 73*, 4293−4298.
(63) Fogolari, F., Corazza, A., Yarra, V., Jalaru, A., Viglino, P., and Esposito, G. (2012) Bluues: a program for the analysis of the electrostatic properties of proteins based on generalized Born radii. *BMC Bioinf. 13*, S18.
(64) Hewitt, F. C., and Samulski, R. J. (2010) Creating a novel origin of replication through modulating DNA-protein interfaces. *PLoS One 5*, e8850.
(65) Hewitt, F. C., Li, C., Gray, S. J., Cockrell, S., Washburn, M., and Samulski, R. J. (2009) Reducing the risk of adeno-associated virus (AAV) vector mobilization with AAV type 5 vectors. *Journal of virology 83*, 3919−3929.
(66) Franke, D., Kikhney, A. G., and Svergun, D. I. (2012) Automated acquisition and analysis of small angle X-ray scattering data. *Nucl. Instrum. Methods Phys. Res., Sect. A 689*, 52−59.
(67) Kozin, M. B., and Svergun, D. I. (2001) Automated matching of high- and low-resolution structural models. *J. Appl. Crystallogr. 34*, 33−41.
(68) Tria, G., Mertens, H. D. T., Kachala, M., and Svergun, D. I. (2015) Advanced ensemble modelling of flexible macromolecules using X-ray solution scattering. *IUCrJ 2*, 207−217.
(69) Hickman, A. B., and Dyda, F. (2005) Binding and unwinding: SF3 viral helicases. *Curr. Opin. Struct. Biol. 15*, 77−85.
(70) Bardelli, M. e. a. Manuscript in preparation.
(71) Petoukhov, M. V., and Svergun, D. I. (2005) Global rigid body modeling of macromolecular complexes against small-angle scattering data. *Biophys. J. 89*, 1237−1250.
(72) Rambo, R. P., and Tainer, J. A. (2011) Characterizing flexible and intrinsically unstructured biological macromolecules by SAS using the Porod-Debye law. *Biopolymers 95*, 559−571.
(73) Chang, Y. P., Xu, M., Machado, A. C., Yu, X. J., Rohs, R., and Chen, X. S. (2013) Mechanism of origin DNA recognition and assembly of an initiator-helicase complex by SV40 large tumor antigen. *Cell Rep. 3*, 1117−1127.
(74) Schneidman-Duhovny, D., Hammel, M., and Sali, A. (2010) FoXS: a web server for rapid computation and fitting of SAXS profiles. *Nucleic Acids Res. 38*, W540−544.

DOI: 10.1021/acs.biochem.5b00610
Biochemistry 2015, 54, 5907−5919

JVI Journals.ASM.org

Oligomeric Properties of Adeno-Associated Virus Rep68 Reflect Its Multifunctionality

Francisco Zarate-Perez,[a] Jorge Mansilla-Soto,[b] Martino Bardelli,[c] John W. Burgner II,[a] Maria Villamil-Jarauta,[a] Demet Kekilli,[d] Monserrat Samso,[a] R. Michael Linden,[c,e] Carlos R. Escalante[a]

Department of Physiology and Biophysics, Virginia Commonwealth University School of Medicine, Richmond, Virginia, USA[a]; Department of Human Genetics, Memorial Sloan-Kettering Cancer Center, New York, New York, USA[b]; Department of Infectious Diseases, King's College London School of Medicine at Guy's, King's and St. Thomas Hospital, London, United Kingdom[c]; Department of Applied Sciences, University of the West of England, Bristol, United Kingdom[d]; UCL Gene Therapy Consortium, UCL Cancer Institute, University College London, London, United Kingdom[e]

The adeno-associated virus (AAV) encodes four regulatory proteins called Rep. The large AAV Rep proteins Rep68 and Rep78 are essential factors required in almost every step of the viral life cycle. Structurally, they share two domains: a modified version of the AAA$^+$ domain that characterizes the SF3 family of helicases and an N-terminal domain that binds DNA specifically. The combination of these two domains imparts extraordinary multifunctionality to work as initiators of DNA replication and regulators of transcription, in addition to their essential role during site-specific integration. Although most members of the SF3 family form hexameric rings *in vitro*, the oligomeric nature of Rep68 is unclear due to its propensity to aggregate in solution. We report here a comprehensive study to determine the oligomeric character of Rep68 using a combination of methods that includes sedimentation velocity ultracentrifugation, electron microscopy, and hydrodynamic modeling. We have determined that residue Cys151 induces Rep68 to aggregate *in vitro*. We show that Rep68 displays a concentration-dependent dynamic oligomeric behavior characterized by the presence of two populations: one with monomers and dimers in slow equilibrium and a second one consisting of a mixture of multiple-ring structures of seven and eight members. The presence of either ATP or ADP induces formation of larger complexes formed by the stacking of multiple rings. Taken together, our results support the idea of a Rep68 molecule that exhibits the flexible oligomeric behavior needed to perform the wide range of functions occurring during the AAV life cycle.

The adeno-associated virus (AAV) nonstructural protein Rep68 belongs to the superfamily 3 of helicases (SF3), whose main characteristic is their multifunctionality (1). Whereas the role of most helicases is to unwind DNA ahead of the replication fork, SF3 helicases also work as initiators of DNA replication, transcriptional regulators, and motor pumps to pack DNA into empty capsids (2). This multifunctionality is due in part to the presence of a specialized AAA$^+$ domain that is at the core of all DNA transactions performed by these proteins. Their catalytic activities are dependent on the formation of oligomeric complexes, where the ATP binding and hydrolysis occurring at the interface of neighboring subunits drive conformational changes that promote translocation or remodeling of target substrates (3). Functional diversity is achieved through formation of different number of oligomers and the presence of specialized associated domains (4). For instance, the AAA$^+$ core of AAV Rep proteins contains a β-hairpin insertion that is involved in the coupling of ATP hydrolysis to DNA translocation/unwinding, and Rep68 and Rep78 have an origin binding domain (OBD) at the N terminus (5–7). Combination of these structural features allows AAV Rep proteins to play a central role in virtually every step of the viral life cycle, such as DNA replication, transcription regulation of the p5 promoter, and site-specific integration (8–11). The AAV Rep proteins have distinct characteristics that position them apart from other SF3 family members, such as simian virus 40 large T antigen (SV40-LTag) and papillomavirus E1 protein (PV-E1). A case in point is the OBD that contains an endonuclease activity that is required for DNA replication and site-specific integration and is similar to domains from other nonstructural proteins from geminiviruses and bacteriophages that use rolling-circle replication (RCR) to replicate their genome (12). Moreover, the minimal AAA$^+$ helicase domain of Rep proteins is monomeric, in contrast to the equivalent domains of SV40-LTag and PV-E1, which form hexameric rings (13–15). This important structural difference may be a reflection of the diversity of DNA binding sites that Rep proteins recognize and the special mechanisms used to replicate the AAV genome. The AAV single-stranded genome contains palindromic sequences at both ends that fold into three-way junction structures called inverted terminal repeats (ITRs). The start of replication initiates at the 3′ end, which is used as a primer for leading strand synthesis. This initial cycle of replication, which requires unwinding of the 5′-end ITR, leaves a DNA molecule with an ITR hairpin end that needs to be replicated to generate a double-stranded DNA (dsDNA) molecule (16, 17). To complete replication, Rep68/Rep78 bind to a specific site called the Rep binding site (RBS) within the ITR and in a reaction that is ATP dependent unwind and nick DNA, generating a new 3′ end (18). The nature of the Rep-DNA complex formed during this process remains unknown, but because of the variety of DNA substrates generated—from double-stranded DNA during the initial binding to single-stranded DNA after the melting reaction—it may

Received 10 September 2012 **Accepted** 5 November 2012
Published ahead of print 14 November 2012
Address correspondence to Carlos R. Escalante, cescalante@vcu.edu. or R. Michael Linden, michael.linden@kcl.ac.uk.

doi:10.1128/JVI.02441-12

require the assembly and disassembly of Rep-DNA complexes with different stoichiometries (19). For instance, binding to both AAV *ori* and AAV S1 sites may be a stepwise process where initial binding to the RBS site is followed by additional recruitment of more Rep molecules, as has been reported for similar initiator proteins such as SV40-Tag and PV-E1. Therein lies the importance of determining the peculiar oligomeric nature of the apo-Rep68/Rep78 proteins in order to understand their DNA-directed assembly. Recently, we and others have determined that the linker that connects the OBD to the AAA^+ domain is critical for the oligomerization of Rep proteins and may be important for DNA complex formation (20, 21). Our current knowledge of Rep68/Rep68 oligomerization derives mostly from studies of these proteins bound to the AAV *ori* or other DNA targets showing a variety of oligomers from hexameric to double octameric (19, 22). In this study, we have used a multifaceted approach involving size-exclusion chromatography, analytical ultracentrifugation, modeling, plus electron microscopy, and we show below that Rep68 has a dynamic behavior in solution and forms a mixture of monomer, dimers, tetramers, and multiple-ring structures.

MATERIALS AND METHODS

Protein expression and purification. Rep68 wild type (wt), as well as the other mutants proteins used here, was expressed and purified as described before (19). Briefly, the His_6-PreScission protease (PP) cleavage site-Rep68 fusion protein was expressed in *Escherichia coli* BL21(DE3)/pLysS bacteria at 37°C for 3 h in LB medium containing 1 mM IPTG (isopropyl-β-D-thiogalactopyranoside). Cell pellets were lysed in Ni-buffer A (20 mM Tris-HCl [pH 7.9 at 4°C], 500 mM NaCl, 5 mM imidazole, 10% glycerol, 0.2% CHAPS {3-[(3-cholamidopropyl)-dimethylammonio]-1-propanesulfonate}, 1 mM TCEP [Tris(2-carboxyethyl)phosphine]). After five 10-s cycles of sonication, the fusion protein was purified using an Ni column preequilibrated in Ni-buffer A. The protein that eluted was desalted using buffer A and a HiPrep 26/10 desalting column (GE Healthcare). The His-PP tag was removed by PreScission protease treatment using 150 μg PP/mg His-PP-Rep68. After overnight incubation at 4°C, buffer was exchanged using the same desalting column and Ni-buffer A. Subsequent Ni column chromatography using buffer B (the same as buffer A but with 1 M imidazole) was performed to remove the uncleaved fusion protein, and untagged Rep68 was eluted with 30 mM imidazole. Rep68 was finally purified by gel filtration chromatography using a HiLoad Superdex 200 16/60 column (GE Healthcare) and size-exclusion buffer (25 mM Tris-HCl [pH 8.0], 200 mM NaCl, 2 mM TCEP). Rep68 wt and mutant proteins were concentrated (1.0 mg/ml for wt and the Cys mutants and 10 mg/ml for the double mutant of Cys). All proteins were flash-frozen in liquid N_2 and stored at −80°C.

Analytical gel filtration chromatography. Samples of the different versions of Rep68 (200 μl) at 4-mg/ml concentrations were chromatographed on a Superose 6 10/300 GL column (GE Healthcare) with a flow rate of 0.5 ml/min. Size-exclusion buffer was used for all chromatographic analyses. Protein elution was detected from its absorbance at 280 nm. A standard curve for protein molecular masses was generated using the proteins carbonic anhydrase (29 kDa), bovine serum albumin (66 kDa), alcohol dehydrogenase (ADH; 150 kDa), β-amylase (210 kDa), apoferritin (443 kDa), and thyroglobulin (669 kDa) (Sigma-Aldrich).

Cross-linking of Rep68*. The cross-linking reactions for the Rep68 Cys151Ser mutant (Rep68*) were done according to a protocol adapted from that of Packman and Perham (23). The protein concentration was 82 μM (5 mg/ml) in 20 mM phosphate buffer, added with 300 mM NaCl, pH 8.0. A 30-fold molar excess of 100 mM dimethyl pimelimidate dihydrochloride (DMP; MP Biomedicals, LLC) was added to the reaction mixture, and the mixture was incubated at room temperature. The reaction was quenched by addition of Tris (pH 7.5) to a final concentration of 50 mM after 1 h.

Analytical ultracentrifugation. Sedimentation velocity experiments were carried out using a Beckman Optima XL-I analytical ultracentrifuge (Beckman Coulter Inc.) equipped with both four- and eight-hole rotors. Samples were loaded in the cells, using in all cases size-exclusion buffer. Samples were centrifuged in 2-sector carbon-filled Epon centerpieces typically at 25,000 rpm and 20°C. Sectors were loaded with a 420-μl sample volume. Typically, 200 or more scans were collected at 5-min intervals at 25,000 rpm. Concentration profiles were collected using both UV absorption (280 nm) and Rayleigh interference optical systems. Results were analyzed using both the SEDFIT and SEDPHAT programs (24, 25).

DNA binding. Fluorescence anisotropy experiments were performed on an ISS PC1 fluorimeter (ISS, Champaign, IL). The DNA substrate used was a single-stranded oligo$(dT)_{38}$ oligonucleotide modified with a 5′ 6-carboxyfluorescein molecule (Integrated DNA Technologies). Reactions were performed in a volume of 500 μl with a buffer containing 25 mM HEPES (pH 7.0), 100 mM NaCl, and a final concentration of 5 nM DNA. Protein concentrations were in the range of 0 to 2.0 μM. Samples were incubated at 20°C for 30 min prior to measurement. Binding activity was followed at an excitation of 492 nm and an emission of 520 nm. Data analysis was performed as described by Yoon-Robarts et al. (7).

Helicase assay. DNA unwinding was measured using a fluorescence resonance energy transfer (FRET)-based fluorometric assay based on the protocol of Bjornson et al., with some modifications (26). The assay uses a DNA molecule labeled at each strand with a FRET pair. The DNA site consists of an 18-bp duplex region and a 10-nucleotide 3′ tail in the bottom strand. The top strand has been labeled with Iowa-Black Dark quencher (IB) at the 3′ end, and the bottom strand has been labeled with cyanine 5 (Cy5) at the 5′ end. IB quenches the fluorescent intensity of Cy5, and upon DNA unwinding, the increase in fluorescent intensity of Cy5 is measured. Fluorescent stopped-flow kinetics studies were carried out using an ISS PC1 fluorimeter (ISS, Champaign, IL) equipped with a Hi-Tech SFA-20 rapid kinetics stopped-flow mixer (TgK Scientific Limited, United Kingdom). The Rep68 protein and the ATP were mixed in a buffer containing 25 mM HEPES, 50 mM NaCl, pH 7.0. A reaction mix of protein, fluorescent DNA (IB/Cy5-dsDNA), and $MgCl_2$ was placed in one syringe of the device, and the ATP mix was placed in a second. Samples were equilibrated at 20°C for 5 min before mixing. Helicase activities were followed using 30 nM fluorescent DNA and 1 μM protein in each case for 900 s. ATP was kept at a 1 mM final concentration. An excitation of 645 nm (Cy5) and an emission of 670 nm (IB) were used for these experiments with a slit of 1 mm. Data results were fit by nonlinear least-squares, using the Origin (version 5.0) program, to $F_t = F_0[1 - \exp(-a \cdot k)]$, where a is the amplitude of the fluorescence, k is the first order rate constant, and F_t and F_0 are the total and initial fluorescent signals, respectively.

ATPase assay. The ATPase assay was performed using a protocol modified from that of Baginski et al. (27). The conditions of the reactions were as follows: 1 mM ATP and 1 μM protein in the presence of 1 mM $MgCl_2$ were mixed in a final volume of 125 μl in reaction buffer A (25 mM Tris-HCl, pH 8.0, 200 mM NaCl, 2 mM TCEP). The reaction mixture was incubated at 25°C, and aliquots were removed every 15 s. This reaction mixture was stopped with 250 μl of solution MA (0.5% ammonium molybdate, 3% ascorbic acid–0.5 N HCl). After 20 min of incubation in ice water, solution CA (2% sodium citrate tribasic dihydrate, 2% sodium arsenite, 2% acetic acid) was added to the samples. After 20 min of incubation at room temperature, samples were read at 840 nm in a spectrophotometer (Agilent Technologies) with a multiple-cell thermoregulated compartment. The formation of inorganic phosphate was calculated from phosphate standard curves.

Virus production and purification. HEK293T cells maintained in Dulbecco's modified Eagle medium (DMEM; Invitrogen, United Kingdom) supplemented with 10% fetal calf serum (FCS; Invitrogen, United Kingdom) at 37°C and in 5% CO_2 were transfected using polyethylenimine (PEI) with 45 μg of pAV2 plasmid (containing the AAV type 2 [AAV2] genome) or its C151S-mutated version, pMB2, and 135 μg of helper plasmid pXX6 (University of North Carolina Vector Core Facility;

see http://genetherapy.unc.edu/mta.htm or a map and the sequence). Cells were incubated in DMEM containing 2% fetal bovine serum, 100 units/ml of penicillin, 100 μg/ml streptomycin, and 25 mM HEPES. After 72 h, the cells were harvested and lysed in 150 mM NaCl, 50 mM Tris at pH 8.5, followed by three freeze (−80°C)-thaw (37°C water bath) cycles. The crude lysate was treated for 30 min at 37°C with 150 units/ml of Benzonase (Sigma) and cleared by centrifugation. The virus was purified on an iodixanol step gradient as previously described (28). The gradient was formed in clear ultracentrifugation tubes (Beckman) by first adding 2.8 ml of 15% iodixanol (Optiprep density gradient medium; Sigma) in 1 M sodium chloride, 1× TD buffer (1× phosphate-buffered saline, 1 mM $MgCl_2$, 2.5 mM KCl) and then underlying, in succession, 1.88 ml 25% iodixanol in 1× TD buffer containing 12.5 μg/ml phenol red (Gibco, Grand Island, NY), 1.55 ml 40% iodixanol in 1× TD buffer, and 1.55 ml 60% iodixanol containing 12.5 μg/ml phenol red. The cell lysate was then applied on top of the gradient, which was centrifuged at 40,000 rpm for 3 h at 18°C in a Sorvall Discovery 90SE ultracentrifuge using a TH641 rotor. The virus was extracted with an 18-gauge needle from the 40%-60% iodixanol interphase as well as the majority of the 40% iodixanol phase. A Vivaspin20, 100-kDa-cutoff concentrator (Sartorius Stedim, Goettingen, Germany) was used to concentrate the virus in a final volume of 700 μl as well as to exchange the buffer to lactated Ringer's solution (Baxter, Deerfield, IL).

Slot blot. HEK293T cells were transfected with pAV2 or pMB2 with Lipofectamine 2000 (Sigma) and after 24 h infected with adenovirus at a multiplicity of infection of 10. After an additional 48 h, the cells were harvested, lysed in 350 μl of 0.2 M NaOH–10 mM EDTA, and boiled for 15 min. Each sample was loaded in triplicate onto a nylon hybridization membrane (Amersham Biosciences). The membranes were rinsed in 2× SSC (1× SSC is 0.15 M NaCl plus 0.015 M sodium citrate), dried, UV cross-linked, and prehybridized in 0.75× nylon wash solution (NW; 40.6 g Na_2HPO_4, 18.65 g EDTA, 500 g SDS in 3.58 liters of H_2O, pH 7.2) buffer at 65°C. The membranes were hybridized overnight in 0.75 × NW buffer to a radiolabeled probe (labeled using a Prime-It RmT random primer labeling kit from Stratagene, [^{32}P]dCTPs from PerkinElmer) consisting of either a *rep* sequence or an ampicillin sequence. Primers used to amplify the *rep* probe were 5′-AACTGGACCAATGAGAACTTTCC-3′ and 5′-AAAAAGTCTTTGACTTCCTGCTT-3′. To generate the ampicillin probe, a 587-bp fragment of the mini-pDG plasmid was amplified by PCR, using the primers ND44 5′-AATCAGTGAGGCACCTATCTCAGC-3′ and 5′-AACTCGGTCGCCGCATACACTATT-3′. The membranes were washed twice with 0.5× NW buffer, followed by an additional wash in 0.1× NW buffer. Finally, the membranes were exposed to a phosphorimager screen for 2 h, and the image was analyzed with ImageQuant software (GE Healthcare Life Sciences).

Determination of infectious particle production. HEK293T cells (10^5) were seeded and triple transfected 2 days later with plasmids encoding the recombinant AAV2 (rAAV2)-green fluorescent protein (GFP) (pTRUF11), AAV Rep (wt or C151S mutant), and Cap (pDG or pMB1) constructs and adenovirus helper functions (pXX6). Cells were harvested 3 days after transfection and treated as described above for virus production. Increasing volumes of crude lysate were used to infect HeLa cells. At 3 days postinfection, HeLa cells were harvested and the percentage of GFP-positive cells was determined by fluorescence-activated cell sorter (FACS) analysis (FACSCalibur; BD).

Sequencing of viral preparations. PCR was used to prepare a fragment of *rep* containing the C151S mutation. The primers used were ND140 (forward, 5′-GTTTCCTGAGTCAGATTCGCG-3′) and ND45 (reverse, 5′-AAAAAGTCTTTGACTTCCTGCTT-3′), both at 0.5pmol/μl. Deoxynucleoside triphosphates (0.2 mM; NEB) and Go *Taq* polymerase enzyme (0.05 U/μl; Promega) were added to the reaction mixture. The cycling parameters were 2 min at 94°C, followed by 35 cycles of 30 s at 94°C, 40 s at 64°C, and 1 min at 72°C, with a final 10 min of incubation at 72°C. The resulting fragment was cloned into the pCR2.1 TOPO vector (Invitrogen) and sequenced using primers 5′-AAAAAGTCTTTGACTTCCTGCTT-3′ and MB3 5′-GTCAGGCTCATAATCTTTCCCGCA-3′ at Eurofins MWG Operon. The results were analyzed using the BLAST (NCBI website) program to confirm the expected viral sequence.

Transmission electron microscopy (TEM) analysis. Protein samples at 0.1 mg/ml were adsorbed directly onto carbon-coated copper grids. Following negative staining with 0.75% (wt/vol) uranyl formate, samples were visualized in an electron microscope (Tecnai F20) operated at 200 kV, and images were collected at a magnification of ×50,000 under low-dose conditions on a Gatan 4k × 4k charge-coupled-device camera. Particle windowing, two-dimensional (2D) alignment, and classification reconstruction were carried out with EMAN2 software. During the entire process, the default settings of this image-processing software were followed for eight iterative alignments. The 2D averages were obtained from a final set of 560 particles.

Generation of Rep68 models. An initial model of AAV2 OBD was built using the crystal structure of AAV5 OBD (Protein Data Bank [PDB] accession number 1RZ9) encompassing residues 1 to 98. The structure of AAV2 Rep40 (PDB accession number 1S9H) contains residues 225 to 490. Using the program MODELLER, we generated the last 46 residues known to be unstructured as an extended polypeptide chain (29). We generated two linker models; the first was modeled as an extended polypeptide. The second model was generated using the Robetta server with the sequence from residues 205 to 230 (30). The three regions (OBD, linker, and helicase domain) were put together using the program COOT (31).

RESULTS

AAV2 Rep68 forms multiple oligomeric species in solution with a tendency to aggregate. An initial characterization of Rep68 in solution was carried out using gel filtration chromatography and sedimentation velocity. Gel filtration profiles on a Superose 6 column clearly showed the presence of three major distributions. A fraction of the protein, P1, eluted at the void volume (V_0), thus indicating the presence of aggregates with molecular masses larger than the exclusion limit of the column, i.e., ~40 MDa (Fig. 1A). The two other species, P2 and P3, eluted at 13.3 and 15.9 ml, which correspond to molecular masses of ~550 kDa and ~110 kDa, respectively. These molecular masses should be considered approximations due to the elongated nature of the Rep68 molecule. At higher concentrations (>1 mg/ml), the fraction of total protein in the aggregate peak increased substantially and constituted the majority of the protein population. This behavior was also observed in sedimentation velocity studies using concentrations ranging from 2 μM to 16 μM (0.125 to 1 mg/ml), indicating that even at the lowest concentration, significant formation of aggregates with an apparent $s_{20,w}$ value of >40S occurs (Fig. 1B). The two other major species have apparent sedimentation coefficients of ~4S and ~13S. The width of the peaks suggests that each of these species may be composed of more than one component. In order to assess this possibility, we analyzed the data using the van Holde-Weischet (VHW) method implemented by the program UltraScan-SOMO (32). This method removes any effect from diffusion on the calculation of $s_{20,w}$, and any spread of these values reflects the presence of multiple species. Figure 1C shows the VHW analysis of the same data, showing significant drifts in both 4S and 13S peaks, indicative of a heterogeneous mixture of multiple oligomeric species.

Oxidation of Rep68 cysteine residues is the major contributor to Rep68 aggregation *in vitro*. Aggregation of Rep68 hinders not only the further characterization of its oligomeric properties but also future functional and structural work. To determine the cause of Rep68 aggregation, we pursued a systematic study on the influence of conditions such as pH, salt, detergents, as well as

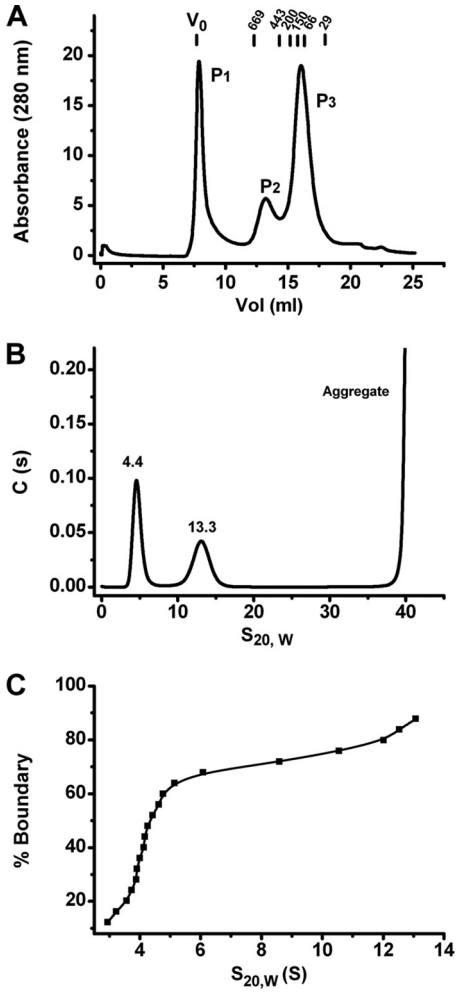


FIG 1 Rep68 wt oligomerization profile. (A) Rep68 wt (2 mg/ml) was injected in a Superose 6 10/300 GL column with a flow rate of 0.5 ml/min. Protein elution was followed by UV detection at 280 nm. The *x* axis represents the elution volume. Molecular mass standards are shown above the plot. (B) Sedimentation velocity profiles of Rep68 wt at a concentration of a 0.5 optical density (0.33 mg/ml). The experiment was run at 40,000 rpm and 20°C, as described in Materials and Methods. (C) van Holde-Weischet analysis of the sedimentation velocity data shown in panel B. The spread from the vertical at both 4S and 13S indicates the presence of more than one species in each population.

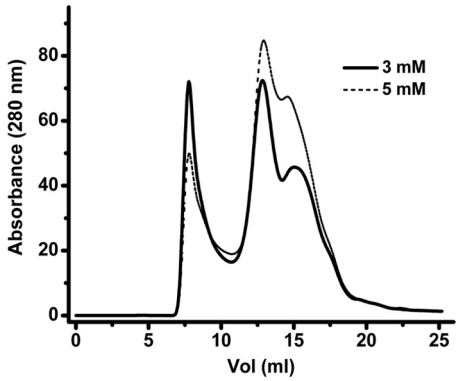


FIG 2 Effect of TCEP on aggregation of Rep68 wt. Rep68 wt (2 mg/ml) was injected with different concentrations of TCEP (3 and 5 mM). A Superose 6 10/300 GL column was used with a flow rate of 0.5 ml/min. Protein elution was followed by UV detection at 280 nm. The aggregation peak of the protein at an elution volume of ~7 ml decreases as the TCEP concentration increases.

chaotropic agents and reducing agents. We determined that only reducing agents and specifically TCEP significantly reduced the fraction of Rep68 aggregates present (Fig. 2). Protein samples at a concentration of 2 mg/ml were analyzed on a Superose 6 column equilibrated with buffer A containing different TCEP concentrations ranging from 3 to 5 mM. These data clearly show that the aggregate peak eluting at the void volume of the column is reduced as the concentration of the reducing agent TCEP is increased. This indicates that aggregation is caused by oxidation, which is perhaps caused by the formation of intermolecular disulfide bonds.

Mutation of Cys151 prevents Rep68 aggregation. To identify cysteine resides with the potential to form intermolecular disulfide bonds, we analyzed the crystal structure of AAV2 Rep40 and an AAV2 OBD model based on the crystal structure of the AAV5 OBD AB (33). The structures show that of the six cysteine residues, four are buried or semiburied in the protein and only two, Cys151 and Cys405, are fully exposed to the solvent. Cys151 resides in the OBD and Cys405 is located in β hairpin 1 of the helicase domain. Interestingly, Cys405 is conserved in other AAV serotypes, while Cys151 is not. We thus proceeded to make single Cys mutants and an additional mutant in which both residues were replaced by serine/alanine. The mutant proteins were purified and analyzed on a Superose 6 column. Figure 3 shows that mutation of Cys151 to alanine or serine abolished the formation of aggregates at all concentrations tested. In contrast, the Cys405Ser mutant protein still showed significant formation of aggregate eluting at the void volume (Fig. 3B). As expected, the double mutant behaved like the single Cys151Ser mutant (Fig. 3C).

The Rep68 C151S mutant is functionally equivalent to Rep68 wt. In order to determine if the Cys151Ser mutation affects any of the biochemical activities of Rep68, we compared the biochemical activities of the C151S mutant with those of Rep68 wt. We performed ATPase (Fig. 4A), helicase (Fig. 4B), and DNA binding functional (Fig. 4C) assays *in vitro*. Results show that the C151S mutant protein performed as well as the wild-type protein in all these functional assays. In addition, we assessed the performance of the C151S mutant in a simple replication assay in HEK293T cells. Figure 4D shows comparable replication efficiencies by both the wt and mutant Rep68 proteins. To determine if the Rep C151S mutant was also able to support the production of infectious viral particles, we produced recombinant AAV2-GFP in the context of the mutant Rep protein and infected HeLa cells with increasing amounts of virus. Figure 4E shows that infectious rAAV2-GFP particles were efficiently formed in the presence of the C151S Rep, albeit at slightly lower titers than wt Rep. To investigate this small difference, we then purified both AAV2 wt and AAV2 containing the C151S mutation and assessed the fraction of empty particles in our preparations by transmission electron microscopy (Fig. 4F). The presence of the mutation in the purified viruses, confirmed by sequencing part of the viral DNA containing the mutation, did not affect the ratio of empty versus full particles.

Rep68* forms multiple species in solution. Similar to the results obtained for Rep68 wt (Fig. 1), Rep68* forms multiple oligomers at moderate protein concentrations. To further character-

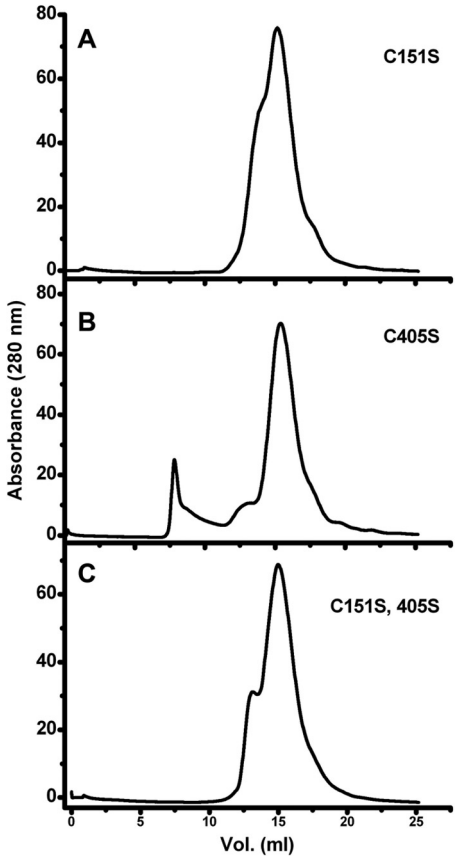


FIG 3 Effect of Cys mutations on Rep68 wt aggregation. Three Rep68 Cys mutants were evaluated by gel filtration chromatography: Rep68 C151S, Rep68 C405S, and a double Cys mutant, Rep68 C151A and C405S. A Superose 6 10/300 GL column was used with a flow rate of 0.5 ml/min. Protein elution was followed by UV detection at 280 nm. When the protein contains the Cys151S mutation (A), no aggregation was observed. In contrast, the C405S mutation still shows an aggregation peak eluting at the void volume of the column (B). (C) The double mutant behaves like the single C151S mutant. The concentration of the proteins was 32.65 μM (2 mg/ml).

ize the oligomeric behavior of Rep68*, an extensive sedimentation velocity study was undertaken at various Rep68* concentrations and in the presence and absence of either ATP or ADP. The results for apo-Rep68* show that the continuous sedimentation [$c(S)$] distributions change significantly as the total concentration of Rep68* is increased (Fig. 5A). At the lowest concentration (1 μM), two species are present, sedimenting at 4S and 13S. The number of species increases at higher concentrations, with as many as six species present at 18 μM and the appearance of a faster-sedimenting species at ~27S. Figure 5B shows the effect of the Rep68* concentration on the weight (actually, the signal) average sedimentation coefficient (s_{wt}) for the overall reaction and the monomer-dimer equilibrium. These plots are considered a sensitive and powerful proof of species at equilibrium (34). The fact that s_{wt} increases significantly over the 18-fold concentration range indicates that dynamic protein-protein interactions are occurring. For the monomer-dimer reaction, s_{wt} was estimated by integrating over a somewhat variable range from 2.5S to 6.5S, which is necessary because at high Rep68* concentrations the larger particles (13S and 27S) interfere with the sedimentation of the monomer and dimer. The pattern exhibited is consistent with a strong monomer interaction that forms dimers (positive slope) and the well-known effect of an increasing concentration of a larger, more rapidly sedimenting species that interferes hydrodynamically. These results, although complex, are consistent with a kinetic process that seems to be near or at the boundary between a slow and an intermediate kinetic process (10^{-4} to 10^{-3}/s and $<10^{-5}$/s, respectively). This is illustrated in the behavior of the 4S peak present at low concentrations but at higher concentrations resolution into two peaks of monomer and dimers as a result of a better signal-to-noise ratio (Fig. 5A). In addition, calculation of the average molecular mass of the 4S species at the lowest concentration gives a molecular mass of ~81 kDa, a value between that of a monomer and that of a dimer. Calculation of the molecular mass of the 13S species gives a range of values from 360 to 450 kDa at the lower concentrations. Together, these results indicate that in solution, Rep68 forms two major populations: a monomer-dimer population that is in slow equilibrium and sediments over the range of 3S to 5S and a 13S population consisting of a mixture of multiple oligomers of six or more Rep68* molecules. Interchange between the two populations occurs as the concentration increases and is mediated by an intermediate species that sediments with an $s_{20,w}$ value of between 5S and 9S, suggesting Rep68* trimers or tetramers. Sedimentation velocity data were also collected in the presence of both 1 mM ATP and ADP, and a representative set of these data at a 18 μM concentration is shown in Fig. 5C. These curves suggest that the presence of either ATP or ADP shifts the equilibrium in favor of higher-order oligomers. This effect can clearly be seen in Fig. 5D, which shows the amount of the 27S peak formed as a function of protein concentration in the presence or absence of nucleotides. The major difference with the apo-Rep68* is the formation of a species sedimenting at ~27S, especially at the lowest concentrations. At the highest concentration (18 μM), two things are occurring: first, we have exceeded the K_d (dissociation constant) for the formation of the 27S species in the absence of nucleotides, and this explains the sudden increase in its formation after 10 μM. Second, the amount of the 27S peak is the same in the absence and presence of nucleotides because we have reached the maximum amount of 27S that can be formed at this concentration. To estimate a value for the dissociation constants of these oligomerizations, each peak was identified by comparing its estimated S value from the peak integrations with S values calculated by hydrodynamic modeling of the structure (see below). Table 1 shows the results of these dissociation constant calculations for the oligomerization steps $K_{2\rightarrow1}$, $K_{4\rightarrow2}$, $K_{8\rightarrow4}$, and $K_{16\rightarrow8}$, where the subscripts represent our best estimation as to the minimum oligomerization state described by the peak under investigation. The one peak that is difficult to assess is that assigned to the tetramer, since the signal from both the trimer and the hexamer might overlap the tetramer peak. The values in Table 1 show a degree of consistency both between values determined for the same step in the presence and absence of nucleotides and between those values calculated for what we presume are linear molecules with the same reactive interface (monomer-dimer and dimer-tetramer). There is a larger ~0.1-fold change in $K_{8\rightarrow4}$ than in $K_{4\rightarrow2}$, which might indicate an entropic effect caused by ring closure. There seems to be a small decrease in the values of the dissociation constants when either ATP or ADP is added; although the effect is not large, the trend seems to be there. The 27S peak must represent a higher-order oligomer, and based on the electron microscopy results pre-

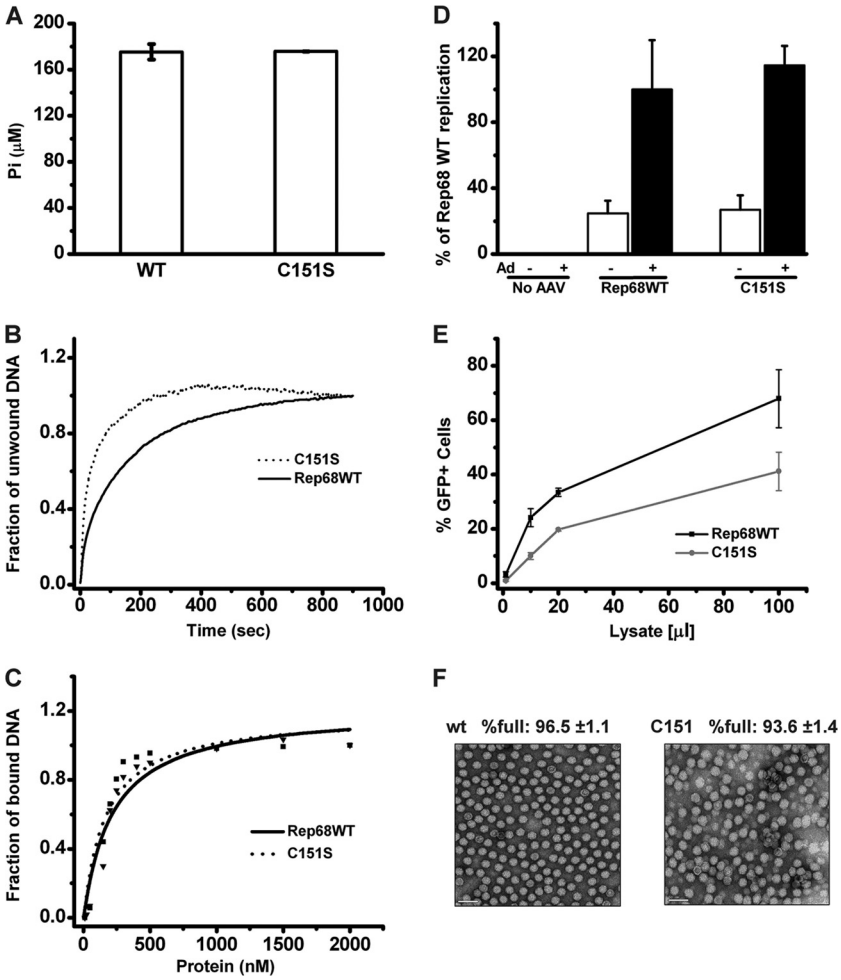


FIG 4 Functional comparison of Rep68 wt and the Rep68 C151S mutant. (A) ATPase activity showed no significant difference for the wt and C151S mutant proteins. (B) Helicase activity for WT and C151S proteins, using fluorescent stopped-flow kinetics). Helicase activities were followed using a 30 nM concentration. (C) Binding of DNA analysis of Rep C151S. A fluorescent probe (F-DNA-dsDNA) was used to monitor the binding of the protein. Binding was followed using 5 nM DNA and a range of protein concentrations from 50 to 2,000 nM. (D) Determination of viral replication in the presence of wt or C151S mutant Rep. The bar graph shows the quantification of three slot blot experiments. The signal obtained with the Rep-specific probe was normalized to that obtained with the ampicillin-specific probe, to normalize for input (transfected) DNA. The replication of the wt AAV2 genome in the presence of adenovirus was set as 100%. Data are presented as means ± SEMs. (E) Comparison of the production of rAAV2-GFP infectious particles in the presence of wt or C151S Rep. rAAV2-GFP particles were produced in HEK293T cells in the presence of adenovirus helper functions, AAV2 Cap, and wt or C151S Rep. Cells were harvested, lysed, freeze-thawed, and treated with the endonuclease Benzonase. Various volumes of crude lysate (in μl, *x* axis) were added to HeLa cells, and the percentage of GFP-positive infected cells was determined by FACS analysis. Data from four experiments are presented as means ± SEMs. (F) Visualization of wt AAV2 and AAV2-RepC151S viral particles by transmission electron microscopy at ×50,000 magnification. Empty viral particles appear as white rings. For each sample, 6 fields of approximately 200 particles were counted; the mean percentage of full particles and the standard deviation are indicated.

sented below, we tend to identify it as a stack set of rings consisting of two either seven- or eight-member rings or a mixture of both.

TEM shows that the 13S population is a mixture of heptameric and octameric rings. In order to determine the exact nature of the 13S species identified in the sedimentation velocity experiments, we carried out negative-stain TEM. We cross-linked all our samples using DMP to stabilize the protein complexes prior to purification on a Superose 6 gel filtration column. All samples were first analyzed by sedimentation velocity to confirm the formation of the 13S complexes. Figure 6A shows a representative electron microscopic image of a negatively stained Rep68*-13S peak where many ring-shaped particles can be seen. A reference-free 2D alignment without imposing symmetry was carried out, and Fig. 6B and C show a selection of several classes, clearly indicating the presence of seven- and eight-ring structures. The heptameric ring has an external diameter of 140.25 Å and an internal diameter of 63.11 Å; the octameric ring has an external diameter of 144 Å and an internal diameter of 67.65 Å. We also analyzed by TEM the effect of ATP on Rep68*. We observed the same seven- and eight-ring structures seen in apo-Rep68*, and in addition, we observed the formation of large filament-like structures. These higher-order structures may represent a series of stacked rings or a helical arrangement of Rep68 molecules, as has been observed in several other AAA^+ proteins and RecA-like proteins (Fig. 6D and E) (35). In addition, we observed the same seven- and eight-ring structures seen in apo-Rep68*.

Hydrodynamic modeling of Rep68 indicates the presence of multiple oligomers. Taking advantage of the X-ray structures of

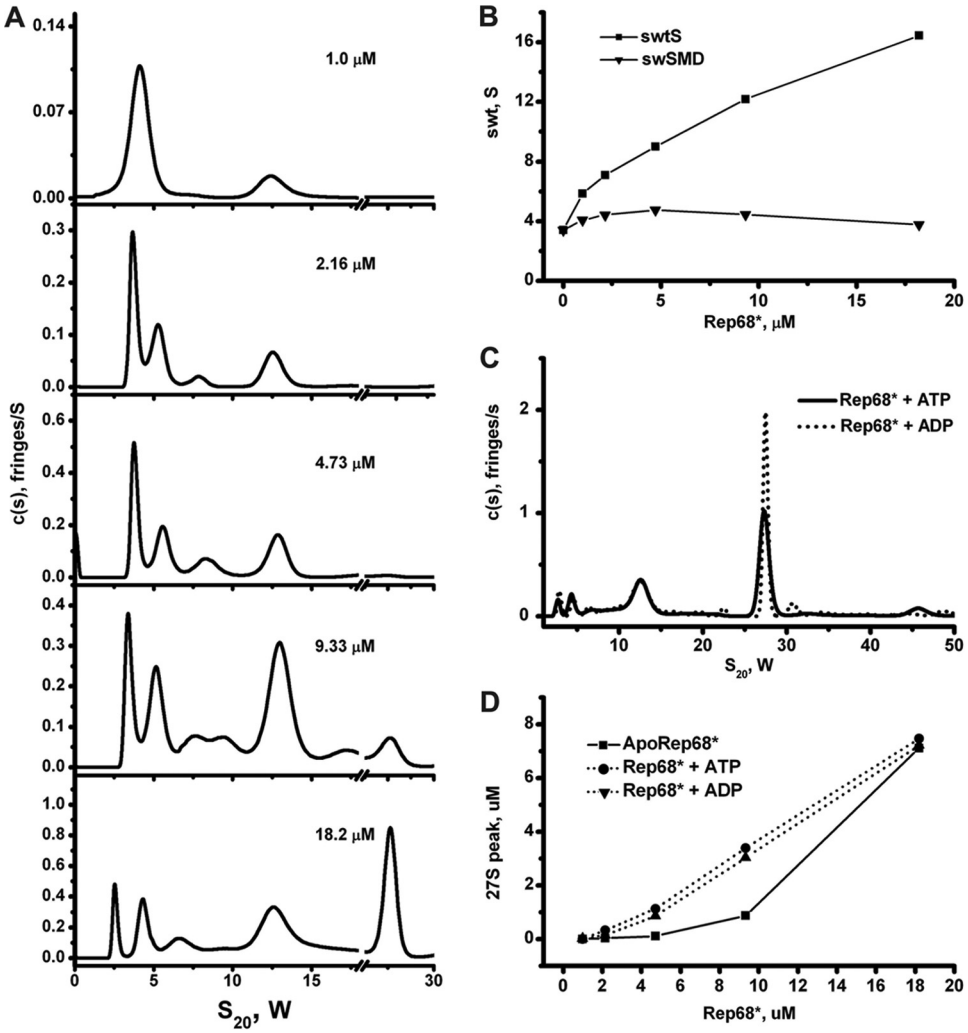


FIG 5 Sedimentation velocity analysis of Rep68*. (A) Sedimentation profiles of Rep68* at five different concentrations obtained using the SEDFIT program. The corresponding values for molecular mass and $s_{20,w}$ are shown in Table 2. (B) Dependence of the weight average sedimentation coefficient on Rep68* loading concentration. ■, weight average sedimentation coefficient for the entire distribution; ▼, weight average for the dimer-monomer equilibrium. (C) Different concentrations of Rep68* (from 1 to 18 μM) were tested in sedimentation analysis assays in the absence or in the presence of nucleotides of ADP and ATP. A final 1 mM concentration of these nucleotides was used in every sedimentation velocity run. (D) Effect of ATP/ADP on the amount of 27S species formed. The area under the 27S peak was integrated at each concentration.

the two functional domains of Rep68, we created several oligomeric models of Rep68 and calculated their hydrodynamic properties using the programs HYDRO and UltraScan-SOMO (36, 37). This approach allows us to verify the results from the sedimentation velocity and TEM experiments. First, in order to examine the limits of our Rep68 model, we generated two structures. The first model has the two domains joined by an extended interdomain linker; in the second model, the linker is folded up. The first structure resulted in a calculated sedimentation coefficient of 2.9S $s_{20,w}$, while the second model predicted a sedimentation coefficient of 3.5S $s_{20,w}$, which is closer to the experimental value. These results suggest that the interdomain linker is not in an extended conformation but probably folded, bringing the two domains close to each other. Using the second model, we generated models for dimers, trimers, and tetramers. At the same time, we created models for hexameric, heptameric, and octameric ring structures. The theoretical sedimentation velocity coefficients of all models were calculated using the program UltraScan-SOMO, as shown in Table 2. The calculated sedimentation coefficients from these models reproduce the experimental values obtained from the sedimentation velocity experiments remarkably well. Taken together, our modeling results support the notion that Rep68 is present as a mixture of monomer-dimer species at slow equilibrium at low concentrations. Increasing the concentration induces the formation of trimers and tetramers that slowly interconvert to several oligomeric rings, namely, heptamers and octamers (Fig. 7).

DISCUSSION

The experiments described here show that *in vitro*, Rep68 exists as a mixture of multiple species in equilibrium. At low concentrations Rep68 partitions into two populations. The first consist of monomers and dimers interchanging slowly and sedimenting as two clearly defined peaks at most concentrations. The second

TABLE 1 Estimates of dissociation constants for the oligomerization of Rep68* at various concentrations and in the presence of 1 mM ATP or ADP[c]

Protein and concn[a] (μM)	K_d (μM)			
	$K_{2\rightarrow1}$	$K_{4\rightarrow2}$	$K_{8\rightarrow4}$	$K_{16\rightarrow8}$
Apo-Rep68*				
1.0	ND[b]	ND	ND	ND
2.1	1.5	1.79	0.16	ND
4.8	3.26	1.71	0.15	ND
9.3	2.35	1.98	0.17	1.6
18.2	1.56	2.56	0.38	0.48
Avg	2.2	2.0	0.21	1.0
Apo-Rep68* with ATP				
0.7	4.9	ND[b]	ND	ND
1.8	2.0	0.43	0.04	0.16
4.2	0.75	0.9	0.13	ND
8.6	0.87	1.26	0.09	0.31
17	0.96	1.1	0.06	0.07
Avg	1.9	0.9	0.08	0.18
Apo-Rep68* with ADP				
0.9	0.9	1.1	0.02	ND
2.1	1.3	0.5	0.04	0.5
4.2	0.7	0.6	0.05	0.4
8.8	0.9	ND	ND	0.7
18.1	1.9	1.0	0.05	0.5
Avg	1.1	0.8	0.04	0.5

[a] Total concentration of the Rep68* monomers in the cell obtained from the integration of the complete $c(S)$ distribution from the interference data.
[b] ND, not determined.
[c] The ~0.1-fold change in $K_{8\rightarrow4}$ relative to that for $K_{4\rightarrow2}$ might indicate a entropic effect caused by ring closure. Note also that there seems to be a general small decrease in the values of the dissociation constants when either ATP or ADP is added; although the effect is not large, the trend seems to be there.

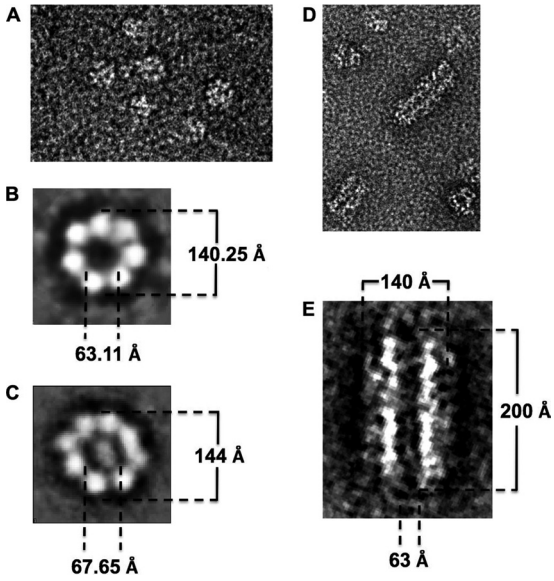


FIG 6 Electron microscopic analysis of the Rep68 protein oligomers. (A) Visualization of cross-linked Rep68* particles by electron microscopy at ×50,000 magnification using the negative-stain technique. Image processing of the particles shows the presence of heptameric (B) and octameric (C) rings. Addition of ATP gives a similar mixture of heptamers and octamers (data not show) but also induces the formation of large filament-like structures (D). (E) Dimension details for these structures after the image processing.

population is a mixture of heptameric and octameric rings that sediment at ~13S. The fact that we can detect the individual peaks of the two populations suggests that they interconvert slower than the time scale of our sedimentation velocity experiments. At higher concentrations, the amount of the 13S population increases and we begin to detect the appearance of an intermediate peak at ~7S. This peak likely represents a mixture of trimers and tetramers. Our data further suggest an oligomerization mechanism where Rep68 heptamers and octamers are assembled through the association of trimers and tetramers from a pool of monomers and dimers (Fig. 7). This process is reversible, as a purified heptamer-octamer mixture dissociates into monomers-dimers upon dilution (data not shown). Similarly, a report by Dignam et al., who used size-exclusion chromatography, showed the concentration-dependent self-association of Rep68 and its tendency to aggregate at moderate salt concentrations, although the nature of the different oligomers was not addressed (38). We have shown here that aggregation of Rep68 is induced by residue Cys151, which has a tendency to form disulfide bridges. *In vivo*, due to the highly reducing environment found in the nucleus, oxidation of cysteines does not occur. The formation of multiple oligomeric rings displayed by Rep68 is similar to the findings for some AAA^+ proteins and other ring helicases. For instance, formation of heptamers has been described for the bacteriophage T7 gene 4 primase/helicase (39), the minichromosome maintenance (MCM) proteins from several species (40–42), the heat shock protein ClpB (43), *Pseudomonas aeruginosa* Hfq (44), ClpP (45), and Rad52 (46). In many cases, mixtures of heptameric and hexameric rings have been observed to be present at the same time (47). Our data suggest that Rep68 adopts a similar behavior, forming heptameric and octameric rings. The functional role of Rep68 ring structures, however, remains to be addressed, and it is pivotal to investigate it in the context of its interaction with different DNA substrates. *In vitro* studies of initiator proteins such as DnaA, SV40-LTag, and PV-E1 show that the formation of an initiation complex is a stepwise process that requires the binding of individ-

TABLE 2 Comparison of experimental and calculated sedimentation coefficients ($s_{20,w}$) for different Rep68 oligomers

Oligomer	Mass (kDa)		$s_{20,w}$ (S)	
	Sequence	Experimental[a]	Experimental[a]	Calculated from model[b]
Monomer	60.8	61.4	3.7	3.4
Dimer	121.6	114	5.4	5.5
Trimer	182.4	215[c]	8.7[c]	7
Tetramer	243.2			8.3
Hexamer	364.8	ND[d]	ND	11.1
Heptamer	425.6	412[c]	13.3[c]	12.5
Octamer	486.4			13.6

[a] Mass and sedimentation coefficients calculated from the corresponding peak position of the sedimentation profiles and f/f_0 values determined with the SEDFIT program.
[b] Sedimentation coefficients calculated from atomic models using the program UltraScan-SOMO.
[c] Calculated from broad peaks that do resolve into two or more peaks with these data.
[d] ND, not determined.

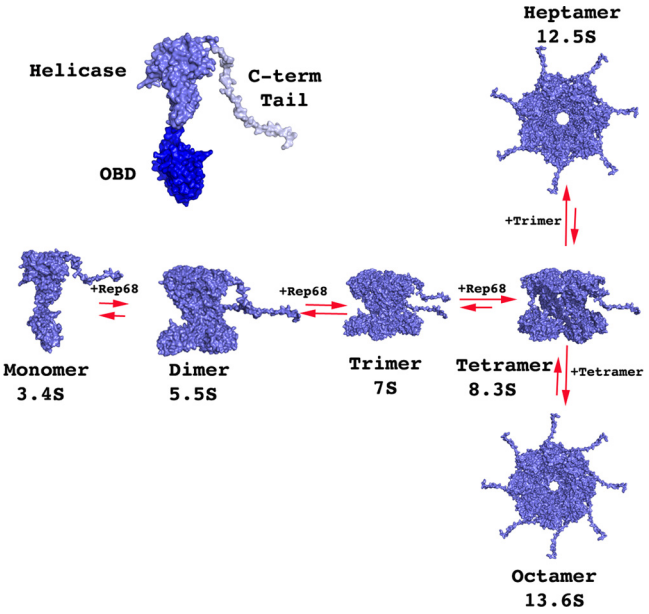


FIG 7 Hydrodynamic modeling of Rep68* oligomerization. A model of full-length Rep68 was generated from the structures of the origin binding domain (dark blue), Rep40 (helicase domain; slate blue), and a flexible C-terminal tail (light blue), as described in Materials and Methods. Models for the different oligomeric rings were generated using the SymmDock server (53). Sedimentation coefficients from each model were calculated using the program UltraScan-SOMO (37).

ual protein molecules to several direct repeats. This process is inhibited if the initiator protein is present as a stable ring structure because of the topological constraints of placing a closed ring around DNA. In many cases, this is what has been observed *in vitro* (13). In addition, formation of rings could provide a more stable structure, increasing protection against degradation. Several groups have postulated that the presence of heptameric rings provides a mechanism for loading the initiator protein onto DNA without the need for an additional loader protein. Thus, upon interaction with DNA, one subunit is lost with the resulting opening of the ring (42, 47). The presence of multiple Rep68 oligomers possibly provides a pool of different functional units that could be used for distinct reactions, such as initiation of DNA replication, recognition of the terminal resolution site (trs) hairpin for nicking, or the binding to the p19 promoter. In this regard, we have shown that Rep68 oligomerization is regulated by DNA structure and found that single-stranded DNA and helicase substrates support the formation of double-octameric Rep68 (19). Moreover, we determined that Rep68 is capable of unwinding a helicase substrate when purified as a double-octamer Rep68-helicase substrate complex. Other reports have found that Rep68 forms hexamers, when bound to dsDNA (22). Another level of complexity that could regulate the oligomerization properties of Rep proteins *in vivo* involves posttranslational modifications. Studies have determined that Rep proteins are highly modified, with residues being phosphorylated, ubiquitinated, and sumoylated. The last two modifications are used primarily to target the Rep proteins for proteosomal degradation (48–50). In contrast, phosphorylation inhibits several enzymatic properties of Rep proteins, such as DNA binding and helicase activity, and may regulate the assembly-disassembly of Rep proteins to DNA substrates (51, 52). The final answer to the functional oligomeric state of Rep68 will be to determine the structure of Rep68 bound to different DNA substrates, such as the AAV *ori*, AAV S1, and p19 promoter and to characterize the effect of phosphorylation on the formation of these complexes.

ACKNOWLEDGMENTS

We thank Rahul Jaiswal, Faik Musayeb, and Soumya G. Remesh for helpful discussions in the preparation of the manuscript. We thank Sergio B. Kaufman and Jeremias Incicco for the ATPase assay protocol.

This work was supported by NIH grant RO1-GM092854 (to C.R.E.) and United Kingdom Medical Research Council grant 1001764 (to R.M.L.).

REFERENCES

1. **Hickman AB, Dyda F.** 2005. Binding and unwinding: SF3 viral helicases. Curr. Opin. Struct. Biol. **15**:77–85.
2. **Singleton MR, Dillingham MS, Wigley DB.** 2007. Structure and mechanism of helicases and nucleic acid translocases. Annu. Rev. Biochem. **76**:23–50.
3. **Erzberger JP, Berger JM.** 2006. Evolutionary relationships and structural mechanisms of AAA$^+$ proteins. Annu. Rev. Biophys. Biomol. Struct. **35**: 93–114.
4. **Ogura T, Wilkinson AJ.** 2001. AAA$^+$ superfamily ATPases: common structure-diverse function. Genes Cells **6**:575–597.
5. **James JA, Escalante CR, Yoon-Robarts M, Edwards TA, Linden RM, Aggarwal AK.** 2003. Crystal structure of the SF3 helicase from adeno-associated virus type 2. Structure **11**:1025–1035.
6. **Owens RA, Weitzman MD, Kyostio SR, Carter BJ.** 1993. Identification of a DNA-binding domain in the amino terminus of adeno-associated virus Rep proteins. J. Virol. **67**:997–1005.
7. **Yoon-Robarts M, Blouin AG, Bleker S, Kleinschmidt JA, Aggarwal AK, Escalante CR, Linden RM.** 2004. Residues within the B′ motif are critical for DNA binding by the superfamily 3 helicase Rep40 of adeno-associated virus type 2. J. Biol. Chem. **279**:50472–50481.
8. **Chiorini JA, Wiener SM, Owens RA, Kyostio SR, Kotin RM, Safer B.** 1994. Sequence requirements for stable binding and function of Rep68 on the adeno-associated virus type 2 inverted terminal repeats. J. Virol. **68**: 7448–7457.
9. **Chiorini JA, Yang L, Safer B, Kotin RM.** 1995. Determination of adeno-associated virus Rep68 and Rep78 binding sites by random sequence oligonucleotide selection. J. Virol. **69**:7334–7338.
10. **Im DS, Muzyczka N.** 1990. The AAV origin binding protein Rep68 is an ATP-dependent site-specific endonuclease with DNA helicase activity. Cell **61**:447–457.
11. **Kyostio SR, Wonderling RS, Owens RA.** 1995. Negative regulation of the adeno-associated virus (AAV) P5 promoter involves both the P5 rep binding site and the consensus ATP-binding motif of the AAV Rep68 protein. J. Virol. **69**:6787–6796.
12. **Iyer LM, Koonin EV, Leipe DD, Aravind L.** 2005. Origin and evolution of the archaeo-eukaryotic primase superfamily and related palm-domain proteins: structural insights and new members. Nucleic Acids Res. **33**: 3875–3896.
13. **Dean FB, Borowiec JA, Eki T, Hurwitz J.** 1992. The simian virus 40 T antigen double hexamer assembles around the DNA at the replication origin. J. Biol. Chem. **267**:14129–14137.
14. **Sedman J, Stenlund A.** 1998. The papillomavirus E1 protein forms a DNA-dependent hexameric complex with ATPase and DNA helicase activities. J. Virol. **72**:6893–6897.
15. **Wessel R, Schweizer J, Stahl H.** 1992. Simian virus 40 T-antigen DNA helicase is a hexamer which forms a binary complex during bidirectional unwinding from the viral origin of DNA replication. J. Virol. **66**:804–815.
16. **Berns KI.** 1990. Parvovirus replication. Microbiol. Rev. **54**:316–329.
17. **Zhou X, Zolotukhin I, Im DS, Muzyczka N.** 1999. Biochemical characterization of adeno-associated virus rep68 DNA helicase and ATPase activities. J. Virol. **73**:1580–1590.
18. **Brister JR, Muzyczka N.** 2000. Mechanism of Rep-mediated adeno-associated virus origin nicking. J. Virol. **74**:7762–7771.
19. **Mansilla-Soto J, Yoon-Robarts M, Rice WJ, Arya S, Escalante CR, Linden RM.** 2009. DNA structure modulates the oligomerization prop-

erties of the AAV initiator protein Rep68. PLoS Pathog. **5**:e1000513. doi:10.1371/journal.ppat.1000513.
20. **Maggin JE, James JA, Chappie JS, Dyda F, Hickman AB.** 2012. The amino acid linker between the endonuclease and helicase domains of adeno-associated virus type 5 Rep plays a critical role in DNA-dependent oligomerization. J. Virol. **86**:3337–3346.
21. **Zarate-Perez F, Bardelli M, Burgner JW, II, Villamil-Jarauta M, Das K, Kekilli D, Mansilla-Soto J, Linden RM, Escalante CR.** 2012. The interdomain linker of AAV-2 Rep68 is an integral part of its oligomerization domain: role of a conserved SF3 helicase residue in oligomerization. PLoS Pathog. **8**:e1002764. doi:10.1371/journal.ppat.1002764.
22. **Smith RH, Spano AJ, Kotin RM.** 1997. The Rep78 gene product of adeno-associated virus (AAV) self-associates to form a hexameric complex in the presence of AAV ori sequences. J. Virol. **71**:4461–4471.
23. **Packman LC, Perham RN.** 1982. Quaternary structure of the pyruvate dehydrogenase multienzyme complex of Bacillus stearothermophilus studied by a new reversible cross-linking procedure with bis(imidoesters). Biochemistry **21**:5171–5175.
24. **Schuck P.** 2003. On the analysis of protein self-association by sedimentation velocity analytical ultracentrifugation. Anal. Biochem. **320**:104–124.
25. **Vistica J, Dam J, Balbo A, Yikilmaz E, Mariuzza RA, Rouault TA, Schuck P.** 2004. Sedimentation equilibrium analysis of protein interactions with global implicit mass conservation constraints and systematic noise decomposition. Anal. Biochem. **326**:234–256.
26. **Bjornson KP, Amaratunga M, Moore KJ, Lohman TM.** 1994. Single-turnover kinetics of helicase-catalyzed DNA unwinding monitored continuously by fluorescence energy transfer. Biochemistry **33**:14306–14316.
27. **Baginski ES, Foa PP, Zak B.** 1967. Microdetermination of inorganic phosphate, phospholipids, and total phosphate in biologic materials. Clin. Chem. **13**:326–332.
28. **Zeltner N, Kohlbrenner E, Clement N, Weber T, Linden RM.** 2010. Near-perfect infectivity of wild-type AAV as benchmark for infectivity of recombinant AAV vectors. Gene Ther. **17**:872–879.
29. **Marti-Renom MA, Stuart AC, Fiser A, Sanchez R, Melo F, Sali A.** 2000. Comparative protein structure modeling of genes and genomes. Annu. Rev. Biophys. Biomol. Struct. **29**:291–325.
30. **Kim DE, Chivian D, Baker D.** 2004. Protein structure prediction and analysis using the Robetta server. Nucleic Acids Res. **32**:W526–W531. doi:10.1093/nar/gkh468.
31. **Emsley P, Lohkamp B, Scott WG, Cowtan K.** 2010. Features and development of Coot. Acta Crystallogr. D Biol. Crystallogr. **66**:486–501.
32. **Brookes E, Demeler B, Rosano C, Rocco M.** 2010. The implementation of SOMO (SOlution MOdeller) in the UltraScan analytical ultracentrifugation data analysis suite: enhanced capabilities allow the reliable hydrodynamic modeling of virtually any kind of biomacromolecule. Eur. Biophys. J. **39**:423–435.
33. **Hickman AB, Ronning DR, Kotin RM, Dyda F.** 2002. Structural unity among viral origin binding proteins: crystal structure of the nuclease domain of adeno-associated virus Rep. Mol. Cell **10**:327–337.
34. **Schuck P.** 2000. Size-distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and Lamm equation modeling. Biophys. J. **78**:1606–1619.
35. **Mueller-Cajar O, Stotz M, Wendler P, Hartl FU, Bracher A, Hayer-Hartl M.** 2011. Structure and function of the AAA^+ protein CbbX, a red-type Rubisco activase. Nature **479**:194–199.
36. **Garcia de la Torre J, Navarro S, Lopez Martinez MC, Diaz FG, Lopez Cascales JJ.** 1994. HYDRO: a computer program for the prediction of hydrodynamic properties of macromolecules. Biophys. J. **67**:530–531.
37. **Rai N, Nollmann M, Spotorno B, Tassara G, Byron O, Rocco M.** 2005. SOMO (SOlution MOdeler) differences between X-ray- and NMR-derived bead models suggest a role for side chain flexibility in protein hydrodynamics. Structure **13**:723–734.
38. **Dignam SS, Correia JJ, Nada SE, Trempe JP, Dignam JD.** 2007. Activation of the ATPase activity of adeno-associated virus Rep68 and Rep78. Biochemistry **46**:6364–6374.
39. **Crampton DJ, Ohi M, Qimron U, Walz T, Richardson CC.** 2006. Oligomeric states of bacteriophage T7 gene 4 primase/helicase. J. Mol. Biol. **360**:667–677.
40. **Costa A, Pape T, van Heel M, Brick P, Patwardhan A, Onesti S.** 2006. Structural basis of the Methanothermobacter thermautotrophicus MCM helicase activity. Nucleic Acids Res. **34**:5829–5838.
41. **Pape T, Meka H, Chen S, Vicentini G, van Heel M, Onesti S.** 2003. Hexameric ring structure of the full-length archaeal MCM protein complex. EMBO Rep. **4**:1079–1083.
42. **Yu X, VanLoock MS, Poplawski A, Kelman Z, Xiang T, Tye BK, Egelman EH.** 2002. The Methanobacterium thermoautotrophicum MCM protein can form heptameric rings. EMBO Rep. **3**:792–797.
43. **Akoev V, Gogol EP, Barnett ME, Zolkiewski M.** 2004. Nucleotide-induced switch in oligomerization of the AAA^+ ATPase ClpB. Protein Sci. **13**:567–574.
44. **Nikulin A, Stolboushkina E, Perederina A, Vassilieva I, Blaesi U, Moll I, Kachalova G, Yokoyama S, Vassylyev D, Garber M, Nikonov S.** 2005. Structure of Pseudomonas aeruginosa Hfq protein. Acta Crystallogr. D Biol. Crystallogr. **61**:141–146.
45. **Thompson MW, Miller J, Maurizi MR, Kempner E.** 1998. Importance of heptameric ring integrity for activity of Escherichia coli ClpP. Eur. J. Biochem. **258**:923–928.
46. **Stasiak AZ, Larquet E, Stasiak A, Muller S, Engel A, Van Dyck E, West SC, Egelman EH.** 2000. The human Rad52 protein exists as a heptameric ring. Curr. Biol. **10**:337–340.
47. **Crampton DJ, Mukherjee S, Richardson CC.** 2006. DNA-induced switch from independent to sequential dTTP hydrolysis in the bacteriophage T7 DNA helicase. Mol. Cell **21**:165–174.
48. **Farris KD, Fasina O, Sukhu L, Li L, Pintel DJ.** 2010. Adeno-associated virus small rep proteins are modified with at least two types of polyubiquitination. J. Virol. **84**:1206–1211.
49. **Sukhu L, Pintel D.** 2011. The large Rep protein of adeno-associated virus type 2 is polyubiquitinated. J. Gen. Virol. **92**:2792–2796.
50. **Weger S, Hammer E, Heilbronn R.** 2004. SUMO-1 modification regulates the protein stability of the large regulatory protein Rep78 of adeno associated virus type 2 (AAV-2). Virology **330**:284–294.
51. **Collaco R, Prasad KM, Trempe JP.** 1997. Phosphorylation of the adeno-associated virus replication proteins. Virology **232**:332–336.
52. **Narasimhan D, Collaco R, Kalman-Maltese V, Trempe JP.** 2002. Hyperphosphorylation of the adeno-associated virus Rep78 protein inhibits terminal repeat binding and helicase activity. Biochim. Biophys. Acta **1576**: 298–305.
53. **Schneidman-Duhovny D, Inbar Y, Nussinov R, Wolfson HJ.** 2005. PatchDock and SymmDock: servers for rigid and symmetric docking. Nucleic Acids Res. **33**:W363–W367. doi:10.1093/nar/gki481.

OPEN ACCESS Freely available online

PLoS PATHOGENS

The Interdomain Linker of AAV-2 Rep68 Is an Integral Part of Its Oligomerization Domain: Role of a Conserved SF3 Helicase Residue in Oligomerization

Francisco Zarate-Perez[1], Martino Bardelli[2], John W. Burgner II [1], Maria Villamil-Jarauta [1], Kanni Das [1], Demet Kekilli[3], Jorge Mansilla-Soto[4], R. Michael Linden[2,5]*, Carlos R. Escalante[1]*

1 Department of Physiology and Biophysics, Virginia Commonwealth University, School of Medicine, Richmond, Virginia, United States of America, **2** Department of Infectious Diseases, King's College London School of Medicine at Guy's, King's and St. Thomas Hospital, London, United Kingdom, **3** Department of Applied Sciences, University of the West of England, Bristol, United Kingdom, **4** Center for Cell Engineering, Department of Human Genetics, Memorial Sloan-Kettering Cancer Center, New York, New York, United States of America, **5** UCL Gene Therapy Consortium, UCL Cancer Institute, University College London, London, United Kingdom

Abstract

The four Rep proteins of adeno-associated virus (AAV) orchestrate all aspects of its viral life cycle, including transcription regulation, DNA replication, virus assembly, and site-specific integration of the viral genome into the human chromosome 19. All Rep proteins share a central SF3 superfamily helicase domain. In other SF3 members this domain is sufficient to induce oligomerization. However, the helicase domain in AAV Rep proteins (i.e. Rep40/Rep52) as shown by its monomeric characteristic, is not able to mediate stable oligomerization. This observation led us to hypothesize the existence of an as yet undefined structural determinant that regulates Rep oligomerization. In this document, we described a detailed structural comparison between the helicase domains of AAV-2 Rep proteins and those of the other SF3 members. This analysis shows a major structural difference residing in the small oligomerization sub-domain (OD) of Rep helicase domain. In addition, secondary structure prediction of the linker connecting the helicase domain to the origin-binding domain (OBD) indicates the potential to form α-helices. We demonstrate that mutant Rep40 constructs containing different lengths of the linker are able to form dimers, and in the presence of ATP/ADP, larger oligomers. We further identified an aromatic linker residue (Y224) that is critical for oligomerization, establishing it as a conserved signature motif in SF3 helicases. Mutation of this residue critically affects oligomerization as well as completely abolishes the ability to produce infectious virus. Taken together, our data support a model where the linker residues preceding the helicase domain fold into an α-helix that becomes an integral part of the helicase domain and is critical for the oligomerization and function of Rep68/78 proteins through cooperative interaction with the OBD and helicase domains.

Citation: Zarate-Perez F, Bardelli M, Burgner JW II, Villamil-Jarauta M, Das K, et al. (2012) The Interdomain Linker of AAV-2 Rep68 Is an Integral Part of Its Oligomerization Domain: Role of a Conserved SF3 Helicase Residue in Oligomerization. PLoS Pathog 8(6): e1002764. doi:10.1371/journal.ppat.1002764

Editor: Craig Meyers, Penn State University School of Medicine, United States of America

Received February 1, 2012; **Accepted** May 3, 2012; **Published** June 14, 2012

Funding: This research was supported by NIH grants RO1-GM092854 (CRE) and the UK Medical Research Council grant 1001764 (RML). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: cescalante@vcu.edu (CRE); michael.linden@kcl.ac.uk (ML)

Introduction

The four adeno-associated virus (AAV) Rep proteins are generated from a single open reading frame by the transcriptional use of two different promoters (p5 and p19) and subsequent alternative splicing mechanisms [1,2,3]. These reactions produce proteins that share three functional domains: an origin binding domain (OBD), a SF3 helicase domain and a putative zinc-finger domain [4,5]. The combination of these domains imparts these proteins with striking multifunctionality. In particular, the larger proteins Rep78 and Rep68 function as initiators of DNA replication, transcriptional regulators, DNA helicases and as key factors in site-specific integration [6]. The smaller Rep proteins Rep40 and Rep52, play a critical role during packaging of viral DNA into preformed empty capsids, where they are thought to be part of the packaging motor complex [7,8,9]. Although in terms of domain architecture the AAV Rep proteins resemble other members of the SF3 protein family, the peculiar OBD with its additional nuclease activity and the complex character of their oligomeric properties, set them apart from other SF3 helicases such as simian virus 40 large T antigen (SV40-LTag) and papilloma virus E1 (PV-E1) proteins [10,11,12,13]. In both of these proteins, the minimal SF3 helicase domain assembles into a hexameric ring in a process that can be induced by the presence of ATP and/or single-stranded DNA [14,15]. In contrast, Rep40 containing only the helicase domain and Rep52 with an additional Zn-finger domain, appear to be monomeric [16,17]. This indicates that oligomerization of AAV Rep proteins requires the presence of both the OBD domain and the helicase domain. This combination imparts both Rep68 and Rep78 with a complex and dynamic oligomeric behavior *in-vitro* that is modulated in large part by the nature of the DNA substrate [18]. The monomeric behavior of both Rep40 and Rep52 is striking in that they appear to contain the required structural features that are present in other SF3 helicase members. The X-ray structures of both SV40-LTag and PV-E1 show that their helicase domains assemble as hexameric rings and that the oligomerization interface is bipartite [15,19]. One interface is formed by the interaction of neighbouring N-terminal

Author Summary

Viruses have to optimize the limited size of their genomes in order to generate the proteins required for infection and replication. Several mechanisms are used to accomplish this including the use of multiple promoters and alternative splicing. These processes generate gene products with diverse functions through the combinatorial assembly of a small number of protein domains. The small genome of the adeno-associated virus has two major open reading frames that generate seven proteins, four non-structural Rep proteins and three capsid proteins. The non-structural Rep proteins share a motor domain that uses hydrolysis of ATP to generate the conformational changes that drive DNA replication, transcriptional regulation, site-specific integration and the packing of viral genome into capsids. These functions depend upon the oligomerization of Rep proteins on specific DNA sites through the cooperation of the N-terminal origin binding domain and the C-terminal helicase domain. We provide evidence that the linker that connects the two domains is an integral feature of the helicase domain and contains a conserved aromatic residue that is critical for oligomerization. This residue emerges to be a signature motif of SF3 helicases and is also present in a subset of bacterial Rep proteins that support rolling circle replication mechanism.

oligomerization domains (OD). The second interface is formed by the interaction of the C-terminal AAA^+ domains and is further stabilized by the presence of nucleotides [11,15]. In order to understand the structural features that promote AAV Rep oligomerization, we pursued in this study a detailed structural comparison of SF3 helicases. We show that the OD domain in Rep40/52 has been hindered in its ability to oligomerize by the transcriptional use of the p19 promoter. This event generates proteins with a smaller OD domain as compared to other SF3 helicases. More importantly, we show that in the context of Rep68/78 the required oligomerization is supported by the interdomain linker which is directly involved in oligomerization interface and we provide evidence that the tyrosine residue preceding the start of Rep40/52 (Y224) is critical in the oligomerization and therefore activity of the large AAV Rep proteins. Taken together, our results support a model where oligomerization of Rep68/78 is mediated by a composite oligomerization interface formed by the OBD, helicase and linker domains, with the latter playing an essential role in the inducing the oligomerization process.

Results

The oligomerization domain (OD) of AAV Rep40 differs from the OD's of other hexameric SF3 helicases

As a first step in our attempt to determine the structural features that promote oligomerization in AAV Rep proteins, we analyzed the oligomeric interface of SF3 family members SV40-LTag and PV-E1. As previously described, the helicase domain contains two subdomains: a N-terminal helical bundle of four α-helices known as the oligomerization domain (OD) and the C-terminal AAA^+ subdomain (Figure 1A). In PV-E1 the oligomerization interface spans both subdomains forming two extended surfaces at opposite faces of the proteins. In the AAA^+ subdomain, one face comprises all the catalytic residues, including: the P-loop, its subsequent helix, the β-strands with the associated Walker B residues, sensor 1 motif, and one side of the β-hairpin (Figure 1B). The neighboring subunit interacts through areas that are located in the α-helices "behind" the β-sheet and on the opposite side of the β-hairpin (Figure 1B). Overall, about 20% of the solvent accessible area takes part in the interface and includes about 34% of all residues. In PV-E1, the OD domain consists of 68 residues forming a four helical bundle. The oligomeric interface comes from interaction of residues located in helices 1 and 4 in one monomer, with residues in helices 2, 3 and part of helix 4 in the other subunit (Figure 1B). Most of the interface is hydrophobic with many tyrosine and isoleucine residues. Similar types of interactions are seen in the interface formed by the SV40-LTag OD domains. This domain is a lot bulkier, spanning 89 residues that form a five-helix bundle. The extra helix originates from an additional Zn-finger motif. Significantly, the OD of Rep40, on the other hand, has only 52 aminoacids and, thus, is significantly shorter than PV-E1 and SV40-LTag OD domains. The direct result of this difference is a decrease in the total accessible surface area by more than 1000 $Å^2$. In addition, the packing of the helices is less compact, producing a more dynamic structure (Figure 1C). We hypothesize that the smaller OD domain of AAV Rep proteins imparts these proteins unique oligomeric properties where the smaller Rep40/52 are mostly monomeric while Rep68/78 -with the additional OBD domain- form oligomers. However, the measurable ATPase activity in all Rep proteins, suggest that Rep40/52 should oligomerize in the presence of nucleotides [20].

AAV-2 Rep40 forms a transient dimer in the presence of nucleotides

To determine if the presence of nucleotides can induce oligomerization of Rep40 -containing the minimal helicase domain-, we carried out sedimentation velocity experiments in the presence and absence of nucleotides at different concentrations. The sedimentation velocity profiles offer a complete characterization of the number and type of oligomers in solution. The data were analyzed using the program sedfit [21,22]. Figure 2A shows plots of the c(s) distribution against the sedimentation coefficient (s) for two concentrations of Rep40 in the absence of nucleotides. A single peak whose $s_{20,w}$ increases slightly with increasing concentrations is observed. The slight but significant increase in s and calculated molar mass is consistent with a weak and transient dimerization (for hydrodynamic reasons, s is expected to decrease with increasing concentrations of an ideal solute). The data where also fitted using the program sedphat to a monomer-dimer association were the process is in rapid exchange on the time scale of the centrifuge [22]. Table 1 shows that the dissociation constant in the absence of nucleotides is $\sim 10^{-3}$ M, which is at the upper end of detection by sedimentation velocity. Similar distributions of Rep40 (at 36 μM) in the presence of either 5 mM ATP or ADP are shown in Figure 2B and 2C. Here an increase is observed in the width of these peaks if compared to those for Rep40 alone. This is a well-understood behavior for a associating system whose exchange kinetics are neither slow of fast on the time scale of the centrifuge, thus, broadening the c(s) distribution peak [23]. The presence of a small shoulder suggest that dimer formation is occurring here as well, although perhaps its rate of dissociation is slower than for Rep40 alone. The s-value of the shoulder is consistent with a transient Rep40 dimer that represents ~0.2% of the total amount of protein. The relatively low ATPase activity of Rep40 reported in the literature supports our model of transient dimerization promoted by the binding and/or hydrolysis of ATP [20].

Addition of linker region to Rep40 constructs induces oligomerization

In order to assess whether the interdomain linker connecting the OBD domain and the helicase domains contains additional

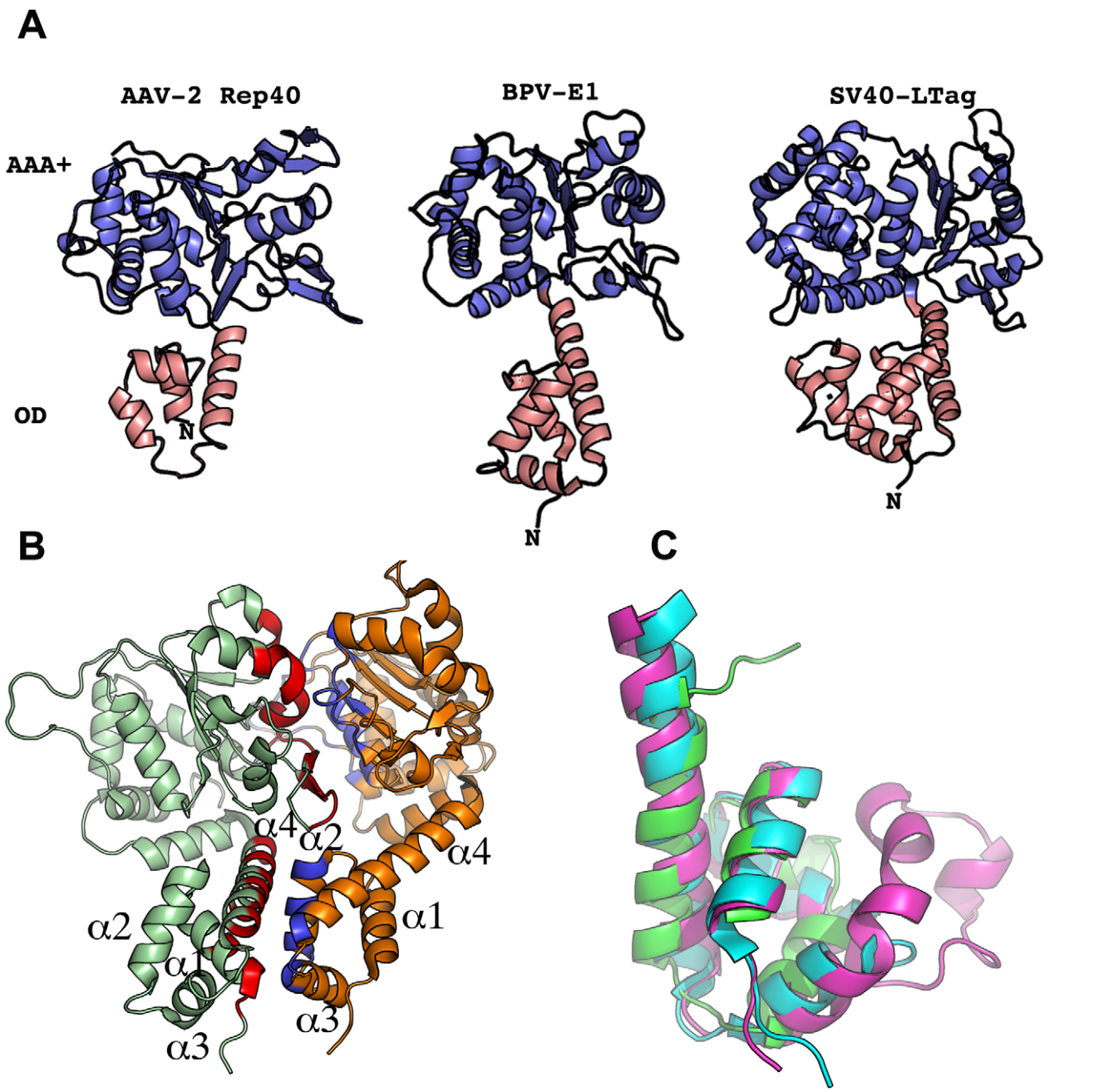


Figure 1. Structural comparison of SF3 helicase structures. (A) Ribbon representation of SF3 helicases AAV-2 Rep40, PV-E1 and SV40-LTag. Salmon color depicts the oligomerization domain (OD). Blue color represents the AAA^{+} domain. (B) PV-E1 dimer showing the residues participating in the formation of the oligomerization interface colored in red and blue. (C) Structural alignment of the OD domain of AAV-2 Rep40 (Green), BPV-E1(Blue) and SV40-LTag(Magenta).
doi:10.1371/journal.ppat.1002764.g001

regions of distinct structure that may play a role in promoting oligomerization, we first carried out secondary structure prediction analysis to determined if the linker contains additional regions of structure. The results suggest that the region from residue 215 to 224 has the potential to form an α-helix (Figure 3A). We hypothesized that this region could extend the first helix of the OD domain (Figure 3A) and the ensuing increase in surface accessible area may be sufficient to drive oligomerization. To test this hypothesis, we designed a new Rep construct beginning at the start of the linker region and extending to aminoacid 536 (a truncated version of Rep68 without the OBD domain, Rep68Δ200), and performed sedimentation velocity and cross-linking studies in order to characterize its oligomerization properties. The sedimentation profile of Rep68ΔN200 shows the presence of two peaks, one corresponding to the monomeric species (~2.53S) and the other to a dimer (~3.71S). The amount of formed dimer increases at higher concentrations as expected from a monomer-dimer equilibrium system (Figure 3B). Formation of dimers was also observed when we performed cross-linking experiments. Figure 3C shows that the amount of dimeric species has significantly

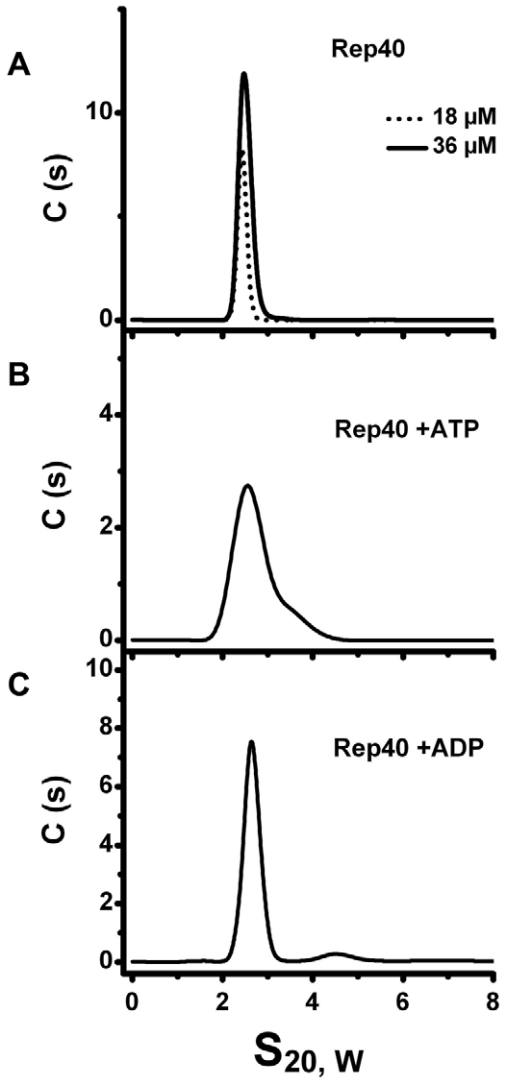


Figure 2. Sedimentation velocity profiles of Re40. (A) Sedimentation velocity analysis of Rep40 at different concentrations (18 μM and 36 μM). In all cases the protein sediments with a sedimentation coefficient of 2.63S. (B) Sedimentation profiles of Rep40 (36 μM) in presence of 5 mM ATP and (C) 5 mM ADP. All the experiments were performed at 40 000 rpm and 20°C in a buffer containing 200 mM NaCl.
doi:10.1371/journal.ppat.1002764.g002

increased in Rep68ΔN200 as compared to Rep40wt. We calculated the dimerization constants of Rep40wt and Rep68ΔN200 from a global fitting of the sedimentation velocity data to a monomer-dimer model (Table 1). In summary, we determined that the presence of the linker region increases the strength of dimerization by about 10-fold relative to that of Rep40.

Extension of the linker region to residue 215 defines the minimal length required to promote oligomerization

Next, we sought to determine the minimal length of linker that is needed to promote oligomerization. We generated three additional constructs, named Rep68ΔN209, Rep68ΔN214 and Rep68ΔN219 and tested their ability to oligomerize (Figure 4). Our results indicate that Rep68ΔN214 contains the minimal length of linker that is required to promote detectible oligomerization, although with the shorter construct Rep68ΔN219, a small shoulder is seen at higher concentration (data not shown). These results confirm that the linker region from 215 to 224 may fold into a α-helix, resulting in an increase of the surface accessible area of the OD domain that mediates oligomerization. This increase, however, is not sufficient to produce higher order oligomers.

Table 1. Global fits of the effect of concentration from sedimentation velocity studies on Rep40 and Rep68ΔN200.

Sample	$s_{monomer}$	s_{dimer}	$logK_a$	K_d	Global χ^2
	S	S		μM	
Rep40	2.63	3.70	3.1	730	0.886
Rep68ΔN200	2.62	4.71	4.1	79	0.20

Sedimentation velocity was performed as described in Materials and Methods. Concentrations of Rep40 used were 18, 36, and 54 μM and for Rep68ΔN200: 4, 8, 16 μM.
doi:10.1371/journal.ppat.1002764.t001

ATP and ADP induce formation of higher order oligomers of the extended linker Rep protein constructs

In order to determine the contribution of ATP and ADP to the oligomerization of the extended linker Rep linker constructs, we performed sedimentation velocity studies in the presence of nucleotides. Our hypothesis was that if oligomerization reflects the functional state of these proteins, the addition of nucleotides should support and induce further oligomerization. Figure 5 shows that the presence of ATP and ADP induces the formation of higher order oligomers. Formation of dimeric species at this concentration can be seen with Rep68Δ214 as well as the longer constructs RepΔN209 and RepΔN200. In the later two, ADP produces two main populations sedimenting at ~3S and ~7S with additional intermediate oligomers. ATP on the other hand, seems to generate more stable species at ~7S. Again, these data show that the presence of the linker region induces oligomerization of the Rep constructs and that the addition of nucleotides, in particular ATP, induces formation of larger oligomers, possibly through the stabilization of the interface formed by the AAA^+ domains. This finding is in good agreement with the unique characteristics of the AAV Rep nucleotide binding pocket, which, based on its open conformation together with the presence of an arginine finger predicts the nucleotide contribution to oligomerization [24].

Linker substitution abolishes oligomerization of Rep68

To determine if the linker is critical for the oligomerization of Rep68, we replaced it with an unrelated sequence and examined its effect on oligomerization using sedimentation velocity. The only prerequisite for the substitute linker were a lack of structure and no impact on the native structures of the connected domains. We chose a sequence from the transcription factor Oct-1. This transcription factor has two DNA binding domains connected by a linker of 29 residues. The X-ray structure of this protein shows that the linker is unstructured and flexible. In addition, it has been used to connect different protein domains without affecting their properties [25,26]. We generated a Rep68 mutant protein ($Rep68_{octlink}$), where residues 206 to 224 were replaced with 18 residues from the Oct-1 linker and tested its ability to oligomerize. The sedimentation profile of Rep68 typically shows two populations with sedimentation coefficients of ~3S and ~13S (Figure 6A). We have determined that the 13S peak corresponds to a mixture of oligomeric rings (data not shown). Figure 6B shows that the replacement of the linker completely abolishes the oligomerization

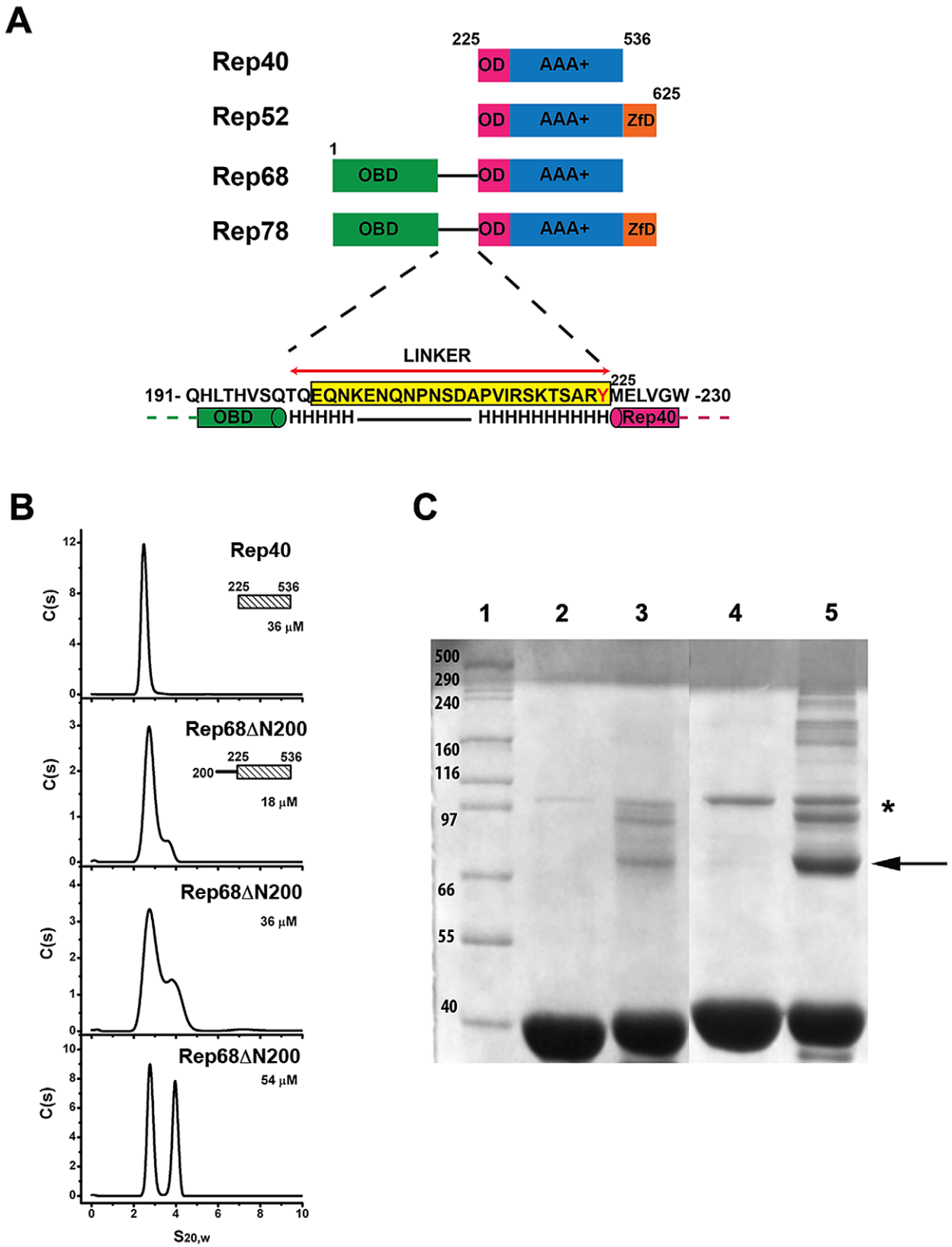


Figure 3. Effect of interdomain linker on oligomerization. (A) A schematic diagram of the AAV-2 Rep proteins with the N-terminal origin binding domain (green), oligomerization domain (pink), AAA^+ domain (blue) and Zinc finger domain (orange). The sequence of the linker region is shown below with the predicted helical regions connecting the OBD and the helicase domains. Linker sequence is highlighted in yellow. (B) Sedimentation profiles of Rep68ΔN200 over the concentration range from 18 to 54 μM and their comparison with Rep40. The sedimentation analysis shows the presence of two peaks of 2.53S and 3.71S, which corresponds to the monomeric and dimeric species respectively. (C) Cross-linking of

Rep40 and Rep68ΔN200. Lane 1, Molecular weight markers; Lane 2, Rep40; Lane3, Cross-linked Rep40; Lane 4, Rep68ΔN200; Lane 5, cross-linked Rep68ΔN200. The arrow shows the position of the Rep68ΔN200 dimer. The asterisk shows proteins bands from non-specific aggregates.
doi:10.1371/journal.ppat.1002764.g003

of the mutant protein $Rep68_{octlink}$. We can detect formation of dimeric species only at the highest concentration tested and in the presence of ATP, (Figure 6C). These results show that replacement of the linker produces a Rep68 protein whose ability to oligomerize has been severely affected.

Presence of the linker region induces oligomerization of the OBD domain

The above findings indicate that the linker region plays a central role in the oligomerization of AAV Rep proteins. To confirm that the linker region has an intrinsic property to induce oligomerization, we generated a construct that spans the OBD domain and the linker region (OBD-linker residues 1–224) and measured its ability to oligomerize. We first analyzed the OBD domain (1–208) to determine any oligomerization up to concentrations of 1 mg/ml (43 μM). Our results show that while OBD is a monomer (Figure 7A), the OBD-linker protein construct displays formation of dimers at increasing protein concentrations (Figure 7B). These results support the hypothesis that the linker region has an intrinsic property to induce oligomerization

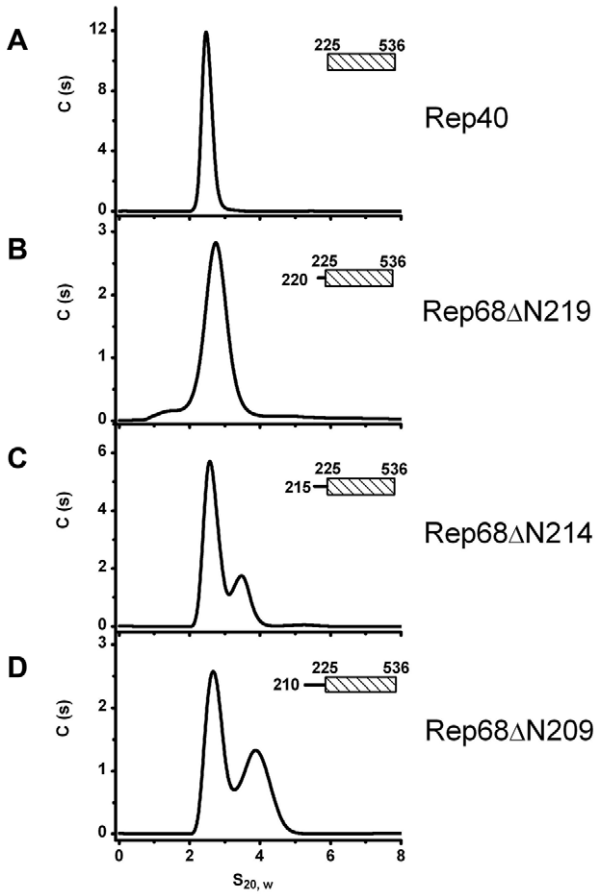


Figure 4. Minimum linker length requirements to induce oligomerization. We generated protein constructs with decreasing lengths of linker that are named in the context of the full Rep68 protein: comparison of the sedimentation velocity profiles of (A) Rep40, (B) Rep68ΔN219 (residues 220–536), (C) Rep68ΔN214 (residues 215–536) and (D) Rep68ΔN209 (residues 210–536). Monomeric species proteins sediment at ~2.7S while the peak at ~3.7S corresponds to a dimer. Protein concentration was at 36 μM in all the sedimentation experiments and run at 40000 rpm and 20°C.
doi:10.1371/journal.ppat.1002764.g004

Linker residue Y224 is critical for oligomerization and represents a conserved feature in SF3 helicases

We generated a model of the Rep68ΔN214 construct using the X-ray structure of Rep40 (residues 225–490) and 9 residues of the linker (215–224) that were added as a helical extension to the N-terminus. The model of the α-helix was generated using Robetta [27]. Figure 8A and 8B shows the structural alignment of the OD domain of the Rep68ΔN214 model with the OD domains of PV-E1 and SV40-LTag. The alignment shows that residue Y224 superimposes with aromatic residues F313 and W270 located at the beginning of helix 1 in the OD domains of PV-E1 and SV40-LTag respectively. Analysis of the structures of both proteins reveals that these aromatic residues play a critical role in forming and stabilizing the oligomerization interface. They pack against both the N-terminal end of helix 4 of the same subunit and the C-terminus end of helix 4 of the neighboring subunit. In order to test the hypothesis that Y224 plays an equivalent role in AAV Rep proteins, we mutated it to alanine and tested its effect on the oligomerization of Rep68ΔN200. Mutation to the smaller residue alanine should have a direct effect in the oligomerization of this protein because of the significant reduction of surface exposed area. Figure 8C shows the sedimentation profile of this mutant protein showing that it completely abolishes the formation of dimers. To confirm that residue Y224 plays an important role in the oligomerization of AAV Rep proteins, we generated a Rep68Y224A mutant and compared its ability to form oligomers with respect to wild type Rep68. Analysis of the Rep68Y224A mutant reveals that at low concentration the protein is mostly found as a monomer with a sedimentation coefficient of ~3S. At higher concentrations, we observed the appearance of multiple peaks that correspond to dimers, trimers and larger oligomers; nevertheless, the majority of the protein is present as a monomer. The presence of ATP induces a small degree of stability to the dimeric species at 5 μM and both the 5S and 11S species at 10 μM. However, the 13S complex observed with the wild type Rep68 is not formed and most of the protein is still found as a monomer (Figure 8E). These results indicate that residue Y224 is critical for the oligomerization of AAV Rep proteins.

Residue Y224 is critical for AAV virus viability

To assess if the disruption of oligomerization observed with the Rep68Y224A mutant has any consequences on the AAV viral life cycle, we produced recombinant AAV2 particles expressing the GFP gene in presence of a helper virus containing the Y224A mutation in the Rep ORF. The cells were harvested and lysed, and the crude lysate (treated with an endonuclease) was used to infect Hela cells. Strikingly, the crude lysate from cells transfected with the mutant helper plasmid didn't contain any infectious rAAV2-GFP particles, as determined by FACS analysis of GFP positive cells (Figure 9). These results show that the residue Y224 of AAV Rep proteins, and the oligomeric properties it confers to these proteins, have a crucial role during the AAV life cycle.

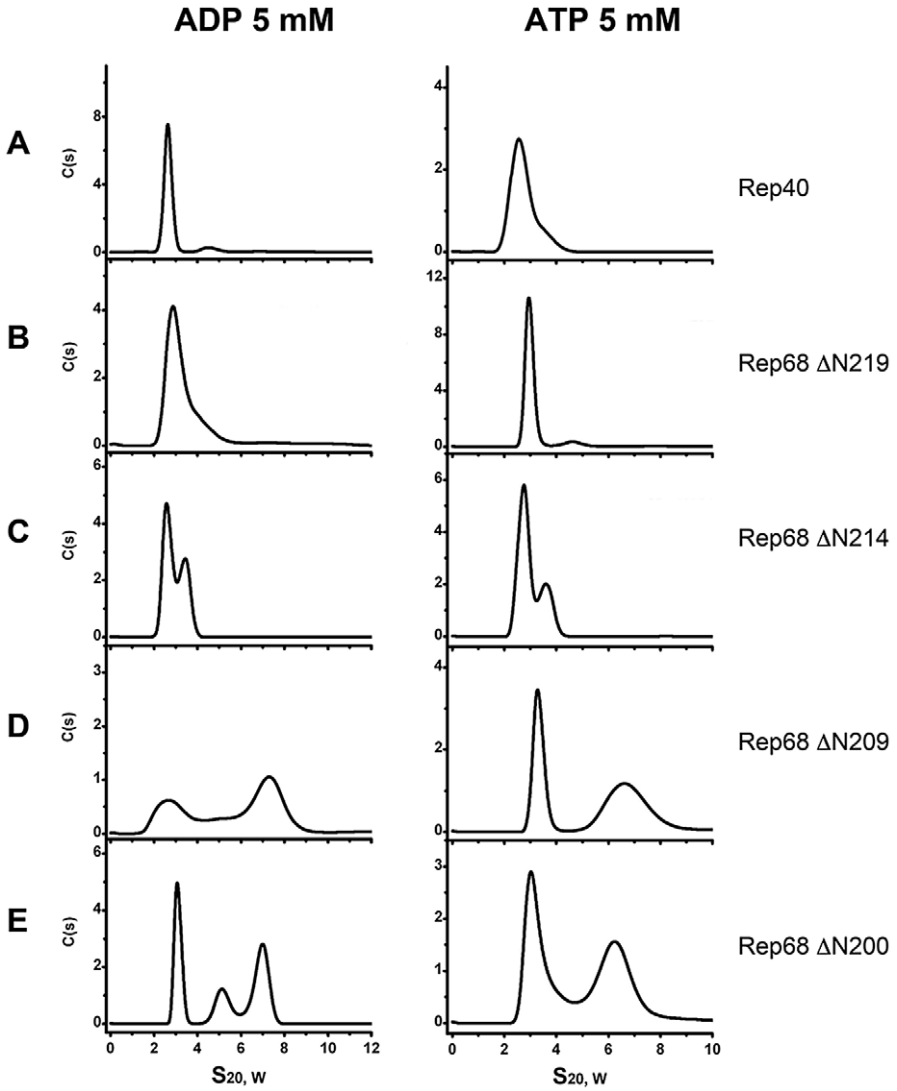


Figure 5. Effect of nucleotides in the oligomerization of Rep extended-linker construct proteins. ATP and ADP were added to each linker construct and compared to Rep40. Sedimentation profiles of (A) Rep40, (B) Rep68Δ219, (C) Rep68Δ214, (D) Rep68Δ209 and (E) Rep68Δ200. All protein concentrations were kept at 36 μM and contain 5 mM of ADP (left panel) or 5 mM ATP (right panel). Sedimentation velocity experiments were run at 40000 rpm and 20°C. Data was collected using the interference system.
doi:10.1371/journal.ppat.1002764.g005

Discussion

In this study we report that the interdomain linker present in the larger AAV Rep68/78 proteins is an integral part of their oligomerization interface. We showed that the linker region is in fact an extension of the OD domain of AAV Rep proteins. Our results have shown that Rep40 constructs containing either a complete or half linker have the ability to oligomerize. This effect is enhanced in presence of ATP or ADP. We hypothesized that the linker region from residues 215 to 224 forms a α-helix that is connected to the first α-helix of the SF3 helicase domain. Secondary structure prediction and modeling of the linker region supports this argument (Figure 3A and 8B). Furthermore, we have identified a critical aromatic residue (Y224) located at the end of the linker region that is conserved in Rep proteins from all AAV serotypes. The bulky nature of this aromatic residue appears to be a conserved feature in SF3 helicases (Figure 8A). Structural alignment of the OD domain of a Rep40 model with an extended helical linker and those of SV40-LTag and PV-E1 shows that residue Y224 aligns with equivalent aromatic residues Trp270 and Phe313 respectively (Figure 8A, 8B). A detailed analysis of the oligomeric interface of these proteins shows that these aromatic residues have a dual role: they stabilize the hydrophobic core of the OD domain helical bundle, and are part of the oligomerization interface between neighboring subunits. Our results reveal the critical role of the OD domain in the formation of stable oligomers in SF3 helicases. The larger OD domains of SV40-Tag and PV-E1 proteins in cooperation with the AAA^+ motor domain generate a helicase domain that forms stable hexamers. Constructs of SV40-LTag and PV-E1 without the OD domain fail to oligomerize [14,19]. Another example that shows the fundamental role of the OD domain in oligomerization comes from the study of the evolutionary related proteins involved in rolling circle replication (RCR) of plasmids. The protein RepB from streptococcal RCR plasmid pMV158 is a hexameric protein that initiates replication of plasmid DNA and has a domain structure that resembles SF3

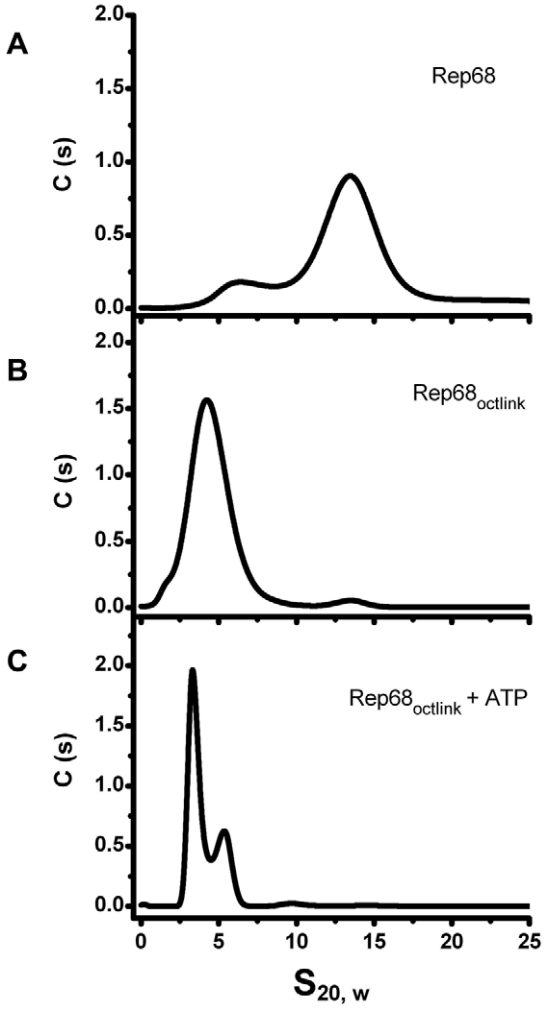


Figure 6. Effect of Linker replacement in Rep68 oligomerization. Comparison of sedimentation profiles of Rep68 and $Rep68_{octlink}$ proteins. (A) Rep68 sediments as a major peak at ~13S. (B) $Rep68_{octlink}$ sediments as a monomer with sedimentation coefficient of 3.5S. (C) $Rep68_{octlink}$ in presence of 5 mM ATP sediments in two peaks corresponding to monomer and dimer species. Protein concentration was kept constant at 25 μM in buffer containing 200 mM NaCl. Sedimentation velocity experiments were run at 40000 rpm and 20°C. Data was collected using the interference system.
doi:10.1371/journal.ppat.1002764.g006

helicases but lacks the AAA^+ subdomain [28]. Its N-terminal OBD domain is structurally and functionally related to the OBD from AAV Rep proteins due to the presence of the HUH motif critical for DNA nicking. Its C-terminal domain only consists of a 4 helical bundle that is similar to the OD domains of SF3 helicases and is responsible for hexamerization. Structural alignment shows that RepB has an aromatic residue (Phe143) equivalent to residue Y224 in AAV Rep68/78. We hypothesize that the role of this residue has been conserved throughout evolution to serve as a modulator of oligomerization in SF3 helicases and related RCR proteins. The smaller AAV Rep proteins Rep40/52 with truncated OD domains are missing the Y224 residue and thus are not able to sustain a stable oligomerization interface and are mostly monomeric. Consequently, the stable oligomerization of AAV Rep proteins requires the cooperative interaction of the OBD domain, the linker and the helicase domain. In this context, the OD sub-domain, and in particular the aromatic residue at the

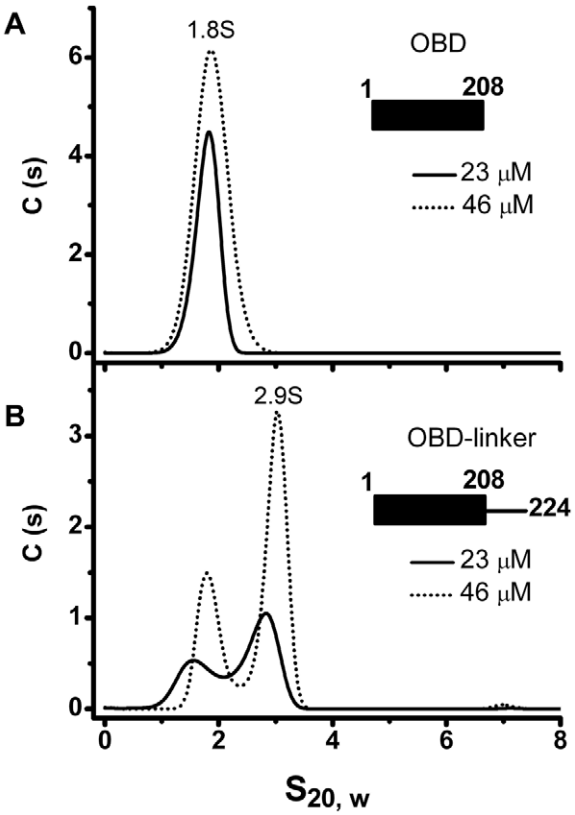


Figure 7. Oligomerization of the OBD domain is induced by the interdomain linker. (A) Sedimentation profile of the OBD (residues 1–208) at two different protein concentrations (23 and 46 μM). (B) Sedimentation profile of OBD-linker (residues 1–224) shows monomers and dimers sedimenting at ~1.8S and ~2.9S respectively. Sedimentation velocity experiments were run at 40000 rpm and 20°C. Data was collected using the interference system.
doi:10.1371/journal.ppat.1002764.g007

C-terminus of linker, appear to be the triggering element required for the oligomerization of AAV Rep proteins.

The critical role of residue Y224 in the overall AAV-2 viral life cycle is illustrated by the complete abolishment of production of infectious particles from AAV-2 vector constructs produced in the context of Rep carrying the Y224A mutation (Figure 9). This result prompts the question of which specific functions are affected by this mutation. We think that most of the biochemical activities of Rep68/78 will be affected due to the impairment in oligomerization. Remarkably, an earlier report by Walker et al. on the identification of residues necessary for site-specific endonuclease activity showed that a Y224 mutant was defective in AAV hairpin/DNA binding, trs endonuclease, DNA helicase and ATPase activity [29], suggesting that correct oligomerization of Rep proteins may be important in all of these functions.

In agreement with our results, a recent report has shown that the presence of the linker in an AAV5 Rep40 construct induces oligomerization in presence of DNA. However, the authors concluded that the linker effect is primarily due to its interaction with DNA [30]. As we demonstrated in this report, the oligomerization effect is an intrinsic property of the linker due to its critical role in the formation of an oligomerization interface as part of the OD domain. The presence of DNA induces further oligomerization as seen with all helicases [13]. However, it appears that the linker also plays an additional role in protein-DNA interaction that may be important during the assembly of Rep68/

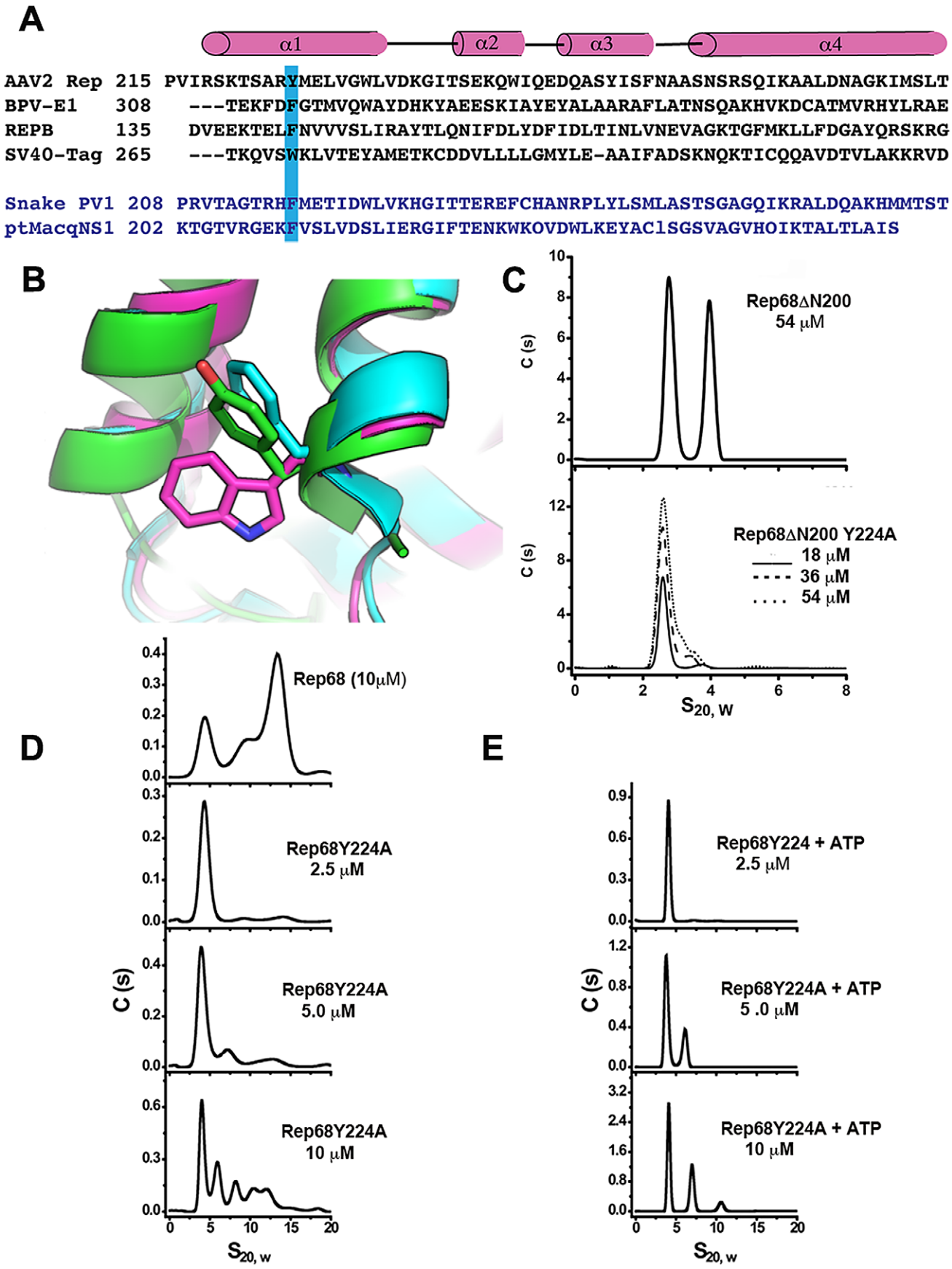

A
α1
α2
α3
α4
AAV2 Rep 215 PVIRSKTSARYMELVGWLVDKGITSEKQWIQEDQASYISFNAASNSRSQIKAALDNAGKIMSLT
BPV-E1 308 ---TEKFDFGTMVQWAYDHKYAEESKIAYEYALAARAFLATNSQAKHVKDCATMVRHYLRAE
REPB 135 DVEEKTELFNVVVSLIRAYTLQNIFDLYDFIDLTINLVNEVAGKTGFMKLLFDGAYQRSKRG
SV40-Tag 265 ---TKQVSWKLVTEYAMETKCDDVLLLLGMYLE-AAIFADSKNQKTICQQAVDTVLAKKRVD
Snake PV1 208 PRVTAGTRHFMETIDWLVKHGITTEREFCHANRPLYLSMLASTSGAGQIKRALDQAKHMMTST
ptMacqNS1 202 KTGTVRGEKFVSLVDSLIERGIFTENKWKOVDWLKEYAClSGSVAGVHOIKTALTLAIS
B
C
Rep68ΔN200
54 μM
C (s)
Rep68ΔN200 Y224A
18 μM
36 μM
54 μM
S20, W
D
Rep68 (10μM)
Rep68Y224A
2.5 μM
Rep68Y224A
5.0 μM
Rep68Y224A
10 μM
C (s)
S20, w
E
Rep68Y224 + ATP
2.5 μM
Rep68Y224A + ATP
5 .0 μM
Rep68Y224A + ATP
10 μM
C (s)
S20, w

Figure 8. Residue Y224 is important for oligomerization. (A) Structure-based sequence alignment of four OD domains: AAV2 Rep, Adeno-Associated virus 2 Rep; BPV-E1, Bovine papillomavirus E1 protein; REPB, Plasmid pMV158 Replication initiator protein; SV40-Tag, Simian virus 40 large T antigen. Also shown in blue the putative OD domains sequences from Snake PV1, Snake parvovirus 1 and ptMacq NS1, Point-tailed Macaque parvovirus non-structural protein 1. Highlighted in blue are the conserved aromatic residues. (B) Ribbon diagram of the structural alignment showing the aromatic residues for AAV-2 Rep (green), SV40-Tag (magenta) and BPV-E1 (Blue). (C) Sedimentation profiles of Rep68ΔN200 and Rep68ΔN200Y224A constructs. (D) Rep68 protein is compared with the mutant Rep68Y224A at different concentrations (2.5 to 10 μM). (E) ATP effect on the oligomerization of Rep68Y224A mutant. Concentration of protein was varied as in figure D from 2.5 μM to 10 μM. ATP and $MgCl_2$ were at 5 mM.
doi:10.1371/journal.ppat.1002764.g008

78 on DNA substrates such as the AAV origin of replication and AAVS1 integration site.

The use of alternative gene promoters is a common mechanism to generate protein diversity and flexibility in gene expression. At the same time it allows to obtain multiple functions from a limited number of genes, thus optimizing the size of the genome. It is clear that in the case of the Rep proteins from the AAV virus, nature has generated two sets of proteins that differ primarily in their ability to oligomerize. Rep proteins obtained from the AAV P_{19} promoter generate Rep40 and Rep52 with truncated OD domains and are thus unable to oligomerize. Both proteins play a critical role during DNA packaging into capsids; however, the mechanism of action of monomeric Rep40/52 during packaging remains elusive. Rep proteins generated from the P_5 promoter, on the other hand, require the cooperative interaction of three different oligomeric interfaces produced by the OBD domain, the linker and the helicase domain. This feature potentially provides an additional dimension for the regulation of the diverse Rep activities when compared to the related proteins from SV40 and PV. We suggest that the cooperative interactions and the modulation of these interfaces – in particular in the presence of various specific DNA substrates – orchestrate the variety of functions performed by Rep68/Rep78 proteins and may thus represent a key to our understanding of the underlying mechanisms.

Finally, our report introduces the possibility of two distinct helicase modes for the biological functions supported by AAV Rep proteins. In the context of the large Rep proteins, a complete OD domain directs the formation of stable oligomers with a DNA unwinding mode likely to resemble that of the related viral proteins SV40-Tag and E1. The small Rep proteins, however, appear to utilize an incomplete OD domain that retains Rep40/52 in a monomeric state with formation of transitional dimeric complexes required for ATP hydrolysis. It is intriguing to speculate that this unique arrangement allows AAV to utilize two distinct motor activities with a single AAA^+ domain. As Rep40/52 have been demonstrated to be required for genome packaging it is feasible to address the question whether this process requires a Rep40/52–mediated dimeric DNA helicase activity by a mechanism that is as yet undiscovered or whether further oligomerization is induced by interaction with capsid proteins.

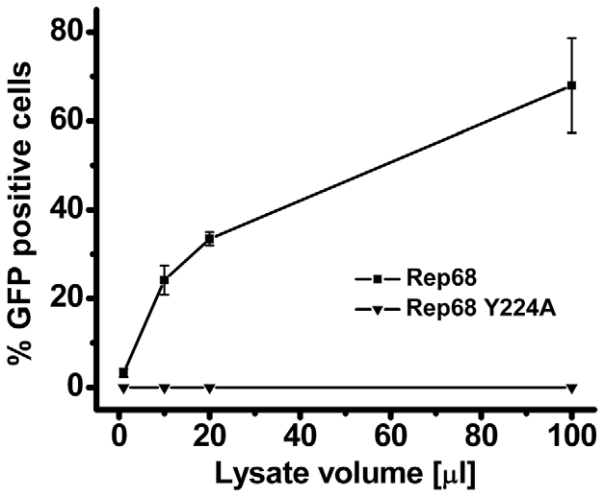


Figure 9. Effect of Y224 mutation on AAV-2 virus liability. Comparison of the production of rAAV2-GFP infectious particles in presence of wt (squares) or Y224A Rep (triangles). rAAV2-GFP particles were produced in 293T cells in presence of wt or Y224A Rep. Varying volumes of crude lysate (in μl, x-axis) were added to HeLa cells and the percentage GFP positive -infected- cells was determined by FACS analysis. Data from four experiments are represented as mean ± s.e.m.
doi:10.1371/journal.ppat.1002764.g009

Materials and Methods

Cloning and mutagenesis of Rep expression constructs

All mutant proteins were generated using the pHisRep68/15b plasmid, which contains the AAV2 Rep68 ORF subcloned in vector PET-15b (Novagen). Site-directed mutagenesis for mutants Y224A was generated using the QuickChange mutagenesis kit (Stratagene). Rep constructs with different linker extensions were generated by PCR with primers designed to encompass the particular protein region. Primers included restriction enzyme sites NdeI and XhoI, and the sequence of the TEV protease site. The Rep68 protein used in these studies contained a Cys to Ser mutation that prevented aggregation but was functionally identical to the wild type protein (data not shown). The $Rep68_{octlink}$ construct was generated by substitution of residues 206 to 224 of AAV2 Rep68 with the mouse Oct-1 linker residues 328–346 (GeneBank CAA49791) using the gene synthesis services from GeneScript. The sequences of all constructs were confirmed by DNA sequencing (GeneWiz).

Protein expression and purification

All proteins were expressed using the pET-15b vector, expressed in E. coli BL21(DE3) cells (Novagen), and purified as described before [18]. The final buffer contains (25 mM Tris-HCl [pH 8.0], 200 mM NaCl, and 2 mM TCEP). His6-PreScission Protease (PP) was expressed in BL21(DE3)-pLysS at 37°C for 3 h, in LB medium containing 1 mM IPTG. Cell pellets were lysed in Ni-Buffer A (20 mM Tris-HCl [pH 7.9 at 4°C], 500 mM NaCl, 5 mM Imidazole, 10% glycerol, 0.2% CHAPS, and 1 mM TCEP). After five 10-s cycles of sonication, the fusion protein was purified using a Ni-column – equilibrated in Ni-buffer A. Protein eluted was desalted using buffer A and a HiPrep^{TM} 26/10 desalting column (GE Healthcare). His-PP tag was removed by PreScission protease treatment using 150 μg PP/mg His-PP-Rep68. After overnight incubation at 4°C, buffer was exchanged using the same desalting column and Ni-Buffer A. Subsequent Ni-column chromatography using the buffer B (same as buffer A but with 1 M imidazole), was performed to remove the uncleaved fusion protein, and untagged Rep68 was eluted with 30 mM imidazole. Rep68 was finally purified by gel filtration chromatography using a HiLoad Superdex 200 16/60 column (GE

Healthcare) and Size Exclusion buffer. N-terminus His6-tagged WT and mutant Rep68 proteins were concentrated to 10 mg/ml, flash-frozen in liquid N_2, and kept at $-80^{\circ}C$ until use.

Cross-linking of Rep40

The cross-linking reactions for Rep40 and Rep68ΔN200 were made according to an adapted protocol from Packman and Perham [31]. The reaction mixture was in cross-linking buffer (25 mM HEPES, 200 mM of NaCl, pH 8.0) and protein concentration was 2 mg/ml. A 30 fold molar excess of 100 mM DMP (dimethyl pimelimidate dihydrochloride, MP Biomedicals, LLC) was added to the reaction and incubated 60 min at room temperature. The reaction was quenched by addition of 1 M Tris, pH 7.5 to a final concentration of 50 mM. The samples were analyzed in an 8% SDS-PAGE.

AAV Infectious particles assay

Hek 293T cells were triple transfected using polyethylenimine (PEI) with an AAV2 ITR-containing plasmid including the GFP gene, a helper plasmid expressing AAV2 Rep (wt or Y224A cloned from the pHisRep68Y224A/15b) and Cap, and a third construct containing the adenovirus helper functions (pXX6, University of North Carolina Vector Core Facility). The presence of the Y224A mutation was confirmed by sequencing (Eurofins). After 72 h, the cells were harvested and lysed in 150 mM NaCl, 50 mM Tris at pH 8.5, followed by three freeze - thaw cycles. The lysate was treated for 30 minutes at $37^{\circ}C$ with 150 units/ml of benzonase endonuclease (Sigma). HeLa cells were infected with increasing amounts of crude lysate, and the percentage of GFP-positive cells was determined three days post-infection.

Analytical ultracentrifugation

Sedimentation velocity experiments were carried out using a Beckman Optima XL-I analytical ultracentrifuge (Beckman Coulter Inc.) equipped with a four and eight-position AN-60Ti rotor. Rep protein samples were loaded in the cells, using in all cases buffer used in the final purification step. Samples in double sector cells were centrifuged at 25,000 rpm for Rep68 proteins (Rep68 and Rep68Y224A). For Rep40 and linker constructs sedimentation was performed at 40,000 rpm. In all experiments, temperature was kept at $20^{\circ}C$. Sedimentation profiles were recorded using UV absorption (280 nm) and interference scanning optics. For the analysis of the results the program Sedfit was used to calculate sedimentation coefficient distribution profiles using the Lamm [21].

Structure analysis and modeling

Structures of AAV-2 Rep40 (1S9H), Bovine papillomavirus E1 protein (2GXA), Simian virus 40 T large antigen (1SVO) and plasmid pMV158 RepB (3DKY) were analyzed using the programs COOT [32], PYMOL [33] and CHIMERA [34]. Structural alignment was done using the DALI server [35]. Secondary structure prediction was performed using PredictProtein [36]. Modeling of the linker region was done using ROBETTA [27].

Acknowledgments

We thank Faik Musayev, Rahul Jaiswal and Soumya G. Remesh for their comments during the writing of the manuscript.

Author Contributions

Conceived and designed the experiments: FZP CRE RML JMS. Performed the experiments: FZP MB CRE JWB MVJ KD DK. Analyzed the data: FZP MB JWB CRE. Wrote the paper: FZP MB RML CRE.

References

1. Green MR, Roeder RG (1980) Transcripts of the adeno associated virus genome: mapping of the major RNAs. J Virol 36: 79–92.
2. Lusby EW, Berns KI (1982) Mapping of the 5′ termini of two adeno-associated virus 2 RNAs in the left half of the genome. J Virol 41: 518–526.
3. Srivastava A, Lusby EW, Berns KI (1983) Nucleotide sequence and organization of the adeno-associated virus 2 genome. J Virol 45: 555–564.
4. Yoon M, Smith DH, Ward P, Medrano FJ, Aggarwal AK, et al. (2001) Amino-terminal domain exchange redirects origin-specific interactions of adeno-associated virus rep78 in vitro. J Virol 75: 3230–3239.
5. Cathomen T, Collete D, Weitzman MD (2000) A chimeric protein containing the N-terminus domain of the adeno-associated virus Rep protein recognizes its target site in an in vivo assay. J Virol 74: 2372–2383.
6. Davis MD, Wu J, Owens RA (2000) Mutational analysis of adeno-associated virus type 2 Rep68 protein endonuclease activity on partially single-stranded substrates. J Virol 74: 2936–2942.
7. Im DS, Muzyczka N (1990) The AAV origin binding protein Rep68 is an ATP-dependent site-specific endonuclease with DNA helicase activity. Cell 61: 447–457.
8. Chiorini JA, Wiener SM, Owens RA, Kyostio SR, Kotin RM, et al. (1994) Sequence requirements for stable binding and function of Rep68 on the adeno-associated virus type 2 inverted terminal repeats. J Virol 68: 7448–7457.
9. Kyostio SR, Wonderling RS, Owens RA (1995) Negative regulation of the adeno-associated virus (AAV) P5 promoter involves both the P5 rep binding site and the consensus ATP-binding motif of the AAV Rep68 protein. J Virol 69: 6787–6796.
10. Hickman AB, Ronning DR, Kotin RM, Dyda F (2002) Structural unity among viral origin binding proteins: crystal structure of the nuclease domain of adeno-associated virus Rep. Mol Cell 10: 327–337.
11. Luo X, Sanford DG, Bullock PA, Bachovchin W (1996) Structure of the origin specific DNA binding domain from simian virus 40 T-antigen. J Virol 3: 1034–1039.
12. Enemark EJ, Chen G, Vaughn DE, Stenlund A, Joshua- Tor L (2000) Crystal structure of the DNA binding domain of the replication initiation protein E1 from papillomavirus. Mol Cell 6: 149–158.
13. Singleton MR, Dillingham MS, Wigley DB (2007) Structure and mechanism of helicases and nucleic acid translocases. Annu Rev Biochem 76: 23–50.
14. Sedman J, Stenlund A (1998) The papillomavirus E1 protein forms a DNA-dependent hexameric complex with ATPase and DNA helicase activities. J Virol 72: 6893–6897.
15. Enermark E, Joshua-Tor L, (2006) Mechanism of DNA translocation in a replicative hexameric helicase. Nature 442: 270–275.
16. Ruiz-Maso JA, Lopez-Zumel C, Menendez M, Espinosa M, del Solar G (2004) Structural features of the initiator of replication protein RepB encoded by the promiscuous plasmid pMV158. Biochim Biophys Acta 1696: 113–119.
17. Smith RH, Kotin RM (1998) The Rep52 gene product of adeno-associated virus is a DNA helicase with 3′-to-5′ polarity. J Virol 72: 4874–4881.
18. Mansilla-Soto J, Yoon-Robarts M, Rice WJ, Arya S, Escalante CR, et al. (2009) DNA structure modulates the oligomerization properties of the AAV initiator protein Rep68. PLoS Pathog 5: e1000513.
19. Dawei L, Zhao R, Lilyestrom W, Gai D, Zhang R, et al. (2003) Structure of the replicative helicase of the oncoprotein SV40 large tumour antigen. Nature 423: 512–518.
20. Dignam SS, Collaco RF, Bieszczad J, Needham P, Trempe JP, et al (2007) Coupled ATP and DNA binding of adeno-associated virus Rep40 helicase. Biochemistry 46: 568–576.
21. Schuck P (2000) Size-distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and lamm equation modeling. Biophys J 78: 1606–1619.
22. Kar SR, Kingsbury JS, Lewis MS, Laue TM, Schuck P (2000) Analysis of transport experiments using pseudo-absorbance data. Anal Biochem 285: 135–142.
23. Correia JJ, Stafford WF (2009) Extracting Equilibrium Constants from Kinetically Limited Reacting Systems. In: Methods Enzymol. Academic Press. 455:419–446.
24. James JA, Aggarwal AK, Linden RM, Escalante CR (2004) Structure of adeno-associated virus type 2 Rep40-ADP complex: insight into nucleotide recognition and catalysis by superfamily 3 helicases. Proc Natl Acad Sci U S A 101: 12455–12460.
25. Klemm JD, Rould MA, Aurora R, Herr W, Pabo CO (1994) Crystal Structure of the Oct-1 POU domain bound to an octamer site: DNA recognition with tethered DNA-binding modules. Cell 77:21–23.
26. Panne D M, T Harrison, S.C. (2004) Crystal Structure of ATF-2/c-Jun and IRF-3 bound to the interferon-B enhancer. EMBO J 23: 4384–4393.

27. Kim DE, Chivian D, Baker D (2004) Protein Structure prediction and analysis using the Robetta server. Nucleic Acids Res 32: W526–531.
28. Boer DR, Ruiz-Maso JA, Lopez-Blanco AG, Vives-Llacer M, Chacon P, et al (2009) Plasmid replication initiator RepB forms a hexamer reminescent of ring helicases and has mobile nuclease domains. EMBO J 28: 1666–1678.
29. Walker SL, Wonderling RS, Owens RA (1997) Mutational Analysis of the Adeno-Associated Virus Rep68 Protein: Identification of Critical Residues Necesssary for Site-Specific Endonuclease Activity. J Virol 71: 2722–2730.
30. Maggin JE, James JA, Chappie JS, Dyda F, Hickman AB (2012) The amino Acid Linker between the Endonuclease and Helicase Domains of Adeno-Associated virus Type 5 Rep Plays a Critical Role in DNA-Dependent Oligomerization. J Virol 86: 3337–3346.
31. Packman LC, Perham RN (1982) Quaternary structure of the pyruvate dehydrogenase multienzyme complex of Bacillus stereарothermophilus studied by a new reversible cross-linking procedure with bis(imidoesters). Biochemistry 21: 5171–5175.
32. Emsley P, Lohkamp B, Scoot WG, Cowtan K (2010) Features and development of Coot. Acta Crystallogr D Biol Crystallogr 66: 486–501.
33. Schrodinger LLC (2010) PyMOL The PyMOL Molecular Graphics System. Version 1.3. Schrodinger,LLC.
34. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, et al (2004) UCSF Chimera - A Visualization System for Exploratory Research and Analysis. J Comput Chem 25: 1605–1612.
35. Holm L, Rosenström P (2010) Dali server: conservation mapping in 3D. Nucleic Acids Res 38: W545–549.
36. Rost B, Yachdav G, Liu J (1997) The PredictProtein Server. Nucleic Acids Res 32: W321–W326.

Identification of a functionally relevant AAV Rep68 oligomeric interface

Martino Bardelli[a], Francisco Zarate-Perez[b], R. Michael Linden[a]*, Carlos R. Escalante[b]#, and Els Henckaerts[a]#

Department of Infectious Diseases, King's College London, London SE1 9RT, UK[a]; Department of Physiology and Biophysics, Virginia Commonwealth University School of Medicine, Richmond, Virginia, USA[b];

Running Head: Critical role of a Rep68 oligomeric interface

#Address correspondence to: Els Henckaerts, els.henckaerts@kcl.ac.uk, or Carlos R. Escalante, carlos.escalante@vcuhealth.org.

*Present address: Genetic Medicine Institute, Pfizer Inc. London, UK

Abstract word count: 245

Total word count: 146

Abstract

The life cycle of the human parvovirus adeno-associated virus (AAV) is orchestrated by four Rep proteins. The large Rep proteins Rep78 and Rep68 are remarkably multi-functional and display a range of biochemical activities, including DNA binding, nicking and unwinding. Functionally, Rep78 and Rep68 are involved in transcriptional

regulation, DNA replication and genomic integration. Structurally, the Rep proteins share a AAA+ domain characteristic of the SF3 family of helicases, with the large Rep proteins additionally containing an N-terminal origin-binding domain (OBD) domain that specifically binds and nicks DNA. The combination of these domains, coupled with dynamic oligomerization properties, is the basis for the remarkable multifunctionality displayed by Rep68 and Rep78 during the AAV life cycle. In this report we describe an oligomeric interface formed by Rep68 and demonstrate how disrupting this interface has drastic effects both on the oligomerization and functionality of the Rep proteins. Our results support a role for the four-helical bundle in the helicase domain of Rep68 as a *bona fide* oligomerization domain (OD). We have identified key residues in the OD that are critical for the stabilization of the Rep68-Rep68 interface; mutation of these key residues disrupts the enzymatic activities of Rep68, including DNA binding and nicking, and compromises viral DNA replication and transcriptional regulation of the viral promoters. Taken together, our data contribute to our understanding of the dynamic and substrate-responsive Rep78/68 oligomerization that is instrumental in the regulation of the DNA transitions that take place during the AAV life cycle.

Importance

The limited genome size of small viruses has driven the evolution of highly multifunctional proteins that integrate different domains and enzymatic activities within a single polypeptide. The Rep68 protein from adeno-associated virus (AAV) combines a DNA binding and endonuclease domain with a helicase-ATPase domain, which together support DNA replication, transcriptional regulation and site-specific

integration. The coordination of the enzymatic activities of Rep68 remains poorly understood, however Rep68 oligomerization and Rep68-DNA interactions have been suggested to play a crucial role. We investigated the determinants of Rep68 oligomerization and identified a hydrophobic interface necessary for Rep68 activity during the AAV life cycle. Our results provide new insights into the molecular mechanisms underlying the regulation of the versatile Rep proteins. Efficient production of AAV-based gene therapy vectors requires optimal Rep expression levels, and studies such as the one presented here could contribute to further optimization of AAV production schemes.

Introduction

Adeno-associated virus (AAV) is a human DNA virus of the family of *parvoviridae* with a unique dependance on helper viruses such as adenovirus or herpes virus for productive replication (1). In order to take advantage of host pathways and helper virus for productive replication with only a limited number of viral gene products at hand, AAV has evolved to combine multiple functions into single proteins. More specifically, a single open reading frame generates the non-structural Rep proteins that orchestrate the different aspects of the AAV life cycle, including transcriptional regulation, replication, packaging and Rep-mediated integration. The four multi-domain Rep proteins, Rep40, Rep52, Rep68, and Rep78, are generated through the use of two promoters and alternative splicing (2, 3). All Rep isoforms share an SF3 helicase domain (HD) that combines ATPase and helicase activities (4, 5). The large Rep proteins Rep68 and Rep78 further contain an N-terminal origin binding domain

(OBD) that specifically binds and nicks the AAV origin (5-7). Rep78 and Rep52 have an additional zinc-finger domain that is involved in interactions with cellular proteins (8, 9). Differences in the domain composition of the Rep isoforms confer specific functionalities to the large and the small Rep proteins. Rep52 and Rep40 are necessary for efficient packaging of the viral DNA into preformed capsids, but are dispensable for viral replication and integration (10). The DNA binding and nicking activities of the OBD of the large Rep proteins on the other hand, are the basis for AAV DNA replication and integration (11, 12). More specifically, Rep78/68 bind to the Rep binding site (RBS) in the inverted terminal repeats (ITRs) and execute a site- and strand-specific nick at the nearby terminal resolution site (*trs*). This process is necessary for resolution of the ITRs and completion of the viral DNA replication cycle with assistance of the host cell machinery (12, 13). Similarly, the DNA binding and endonuclease activities are required for mediation of integration at chromosomal target loci that contain RBS/*trs* sites such as the integration hotspot *AAVS1 (14-16).* In addition*,* efficient nicking of the *trs* at both the viral and cellular origin requires ATP-dependent helicase activity for the generation of an optimal single-stranded substrate (7, 17). Finally, both OBD and helicase domains have the ability to mediate transcriptional regulation of viral and cellular promoters by two independent mechanisms conferring regulatory functions to both small and large Rep proteins (18-20).

Structurally, the Rep proteins belong to the superfamily 3 of helicases (SF3), a group of multifunctional viral proteins combining a characteristic AAA^{+} motor domain, which couples ATP hydrolysis and DNA unwinding, with an origin binding domain to achieve rapid origin melting (21). Other members of this family include the simian

95 virus 40 large T antigen (SV40-LTag) and the papilloma virus E1 (PV-E1). In contrast
96 to the OBD of Rep, SV40-LTag and PV-E1 lack endonuclease activity. The AAV large
97 Rep proteins are also related to HUH endonucleases that catalyse rolling-circle
98 replication (RCR) in bacteriophages and geminiviruses, as well as in bacteria (22, 23),
99 indicating a significant evolutionary conservation. The large AAV Rep proteins have a
100 complex and dynamic oligomeric behaviour that can adapt to the different DNA
101 substrates present during the AAV life cycle, varying from the RBS-containing
102 double-stranded DNA encountered during initial origin binding, to the single-
103 stranded DNA after origin melting (24, 25). In contrast to the AAA^+ domains of SV40-
104 LTag and PV-E1, which readily form hexameric rings (26, 27), Rep40 and Rep52 are
105 monomeric, due to the absence of a complete oligomerization domain (OD) at the N-
106 terminus of the SF3 helicase domain (28). Mutation of the corresponding OD in
107 SV40-LTag and PV-E1 has been shown to prevent their oligomerization and to disrupt
108 the replication of SV40 and PV, respectively (29, 30). Intriguingly, a similar OD is also
109 found in oligomeric HUH endonucleases such as RepB from the pMV158
110 streptococcal RCR plasmid despite the absence of a helicase domain (31). In the large
111 Rep proteins, the linker connecting the OBD and the helicase domain provides the
112 residues necessary to complete the OD and plays a crucial role in the oligomerization
113 of the large Rep proteins (28, 32). Thus effectively the OBD, linker and helicase
114 domain cooperatively interact to promote oligomerization.
115 While recent findings have significantly contributed to the understanding of the
116 determinants of Rep oligomerization (25, 28, 32), its relevance in the context of the
117 AAV life cycle remains to be elucidated. In order to gain an understanding on how
118 oligomerization contributes to the multiple enzymatic functions of the Rep proteins,

we took advantage of our previous findings, which show that the linker domain, and in particular the N-terminal linker residue Y224, is essential for Rep oligomerization (28). This residue is found in an equivalent position as the residues in SV40-LTag and PV-E1 known to be crucial for the formation and maintenance of the oligomeric interface (28). Here we describe an oligomeric interface for Rep, identified in a dimeric complex modelled using the Rep40 structure with a predicted extended N-terminal α-helix to complete the OD (33). This model highlights a potential role for Y224 and I251 in the formation of the oligomeric interface; site-directed mutagenesis confirmed that oligomerization is indeed hampered when Y224 and I251 are altered. Moreover, we could demonstrate that mutations that lead to a disruption of the interface result in defects in DNA binding, *trs* nicking and alters the expression levels of the viral proteins, with severe consequences on viral DNA replication and production of infectious virus.

Materials and Methods

Protein production and purification

All mutations were generated in the pHisRep68/15b plasmid, which contains the AAV2 Rep68 ORF subcloned in the vector pET-15b (Novagen) using the Quickchange™ mutagenesis kit (Agilent Technologies Inc). All proteins were expressed in E.coli BL21(DE3) cells (Novagen), and purified as described previously (28). In brief, cell pellets were lysed in Ni-Buffer A (20 mM Tris-HCl [pH 7.9 at 4 °C], 500 mM NaCl, 5 mM Imidazole, 10% glycerol, 0.2% CHAPS, and 1 mM TCEP) and purified using a Ni-column. The hexa-histidine tag was removed using the PreScission

protease and Rep68 was further purified by gel filtration chromatography using a HiLoad Superdex 200 16/ 60 column (GE Healthcare) and size exclusion buffer (25 mM Tris-HCl [pH 8.0], 200 mM NaCl, and 2mM TCEP). Rep68 WT and mutant proteins were concentrated to 10 mg/ml, flash-frozen in liquid N_2, and kept at -80°C.

Sedimentation velocity experiments

Analytical ultracentrifugation experiments were carried out using a Beckman Optima XL-I analytical ultracentrifuge (Beckman Coulter Inc.) equipped with both four- and eight-position rotors. Protein samples (420µl, 10µM final concentration) were loaded in the cells, using in all cases buffer containing 25mM Tris-HCl, pH 8.0 and 200mM NaCl. Samples were centrifuged in 2-sector carbon-filled Epon centerpieces at 25,000rpm at 20°C. At least 200 scans were collected at 5-min intervals at 25,000 rpm. Sedimentation velocity concentration profiles were collected using both UV absorption (280 nm) and Rayleigh interference scanning optics. Results were analyzed using the SEDFIT program (34, 35).

AAV Infectious particles assay

293T cells were triple-transfected with an AAV2 ITR-containing plasmid encoding a CAG-controlled GFP gene (pTRUF11), a helper plasmid expressing AAV2 Rep (WT or mutants) and Cap, and a third construct containing the adenovirus helper functions (HGTI plasmid) (36, 37). The mutations in Rep were confirmed in all plasmids by sequencing (Eurofins). After 72h, the supernatant was harvested, spun to clear cellular debris and increasing volumes of supernatant were used to transduce HeLa cells. The percentage of GFP-positive HeLa cells was determined 48h post-transduction by flow cytometry (FACSCanto, BD Biosciences).

qPCR-based replication assay

293T cells were transfected with PEI (Polysciences, Inc) and the infectious AAV plasmid pAV2 (38) or its mutant versions, and super-infected 4h later with adenovirus serotype 5 (Ad5) at a multiplicity of infection (MOI) of 5. After 72h, cells were harvested in PBS, pelleted, and the pellet was divided in 4. A fourth was used for RNA extraction, one for protein extraction, one for total DNA extraction and the last quarter for Hirt extraction of small molecular weight DNA.

Total DNA was extracted using the Qiagen DNAeasy Blood and Tissue DNA extraction kit. Viral DNA was quantified by real-time PCR using the SYBR Green JumpStart Taq ReadyMix for qPCR (Sigma-Aldrich) using an ABI PRISM system (Applied Biosystems). Cap primers (fw: TTCTCAGATGCTGCGTACCGGAAA; rv: TCTGCCATTGAGGTGGTACTTGGT) and a pAV2-based standard curve were used for absolute quantification; the signal was normalised to cyclophillin (fw: TGCTGGACCCAACACAAATG; rv: TGCCATCCAACCACTCAGTCT).

Analysis of replicative intermediates

293T cells were treated as described for the qPCR-based replication assay. Small molecular weight was extracted using a modified version of the Hirt extract procedure (39). Briefly, cells were lysed in Hirt Lysis buffer (0.6% SDS, 10mM Tris pH 7.5, 10mM EDTA), and treated with proteinase K (Thermo Fisher) to digest proteins. The high molecular weight DNA was precipitated and discarded. The small molecular weight DNA was then purified by phenol extraction followed by sodium acetate and isopropanol precipitation. The precipitated DNA was washed and resuspended in DNAse-free water. The extracts were digested with the restriction enzyme *Dpn*I

189 (New England Biolabs) to digest input DNA. Samples were run on a 0.8% agarose gel
190 at 30V overnight, and transferred to a nitrocellulose membrane by the Southern
191 blotting method. The membranes were hybridized overnight in 0.75 nylon wash
192 buffer (40.6g Na2HPO4, 18.65g EDTA, 500g SDS in 3.58 liters of ddH2O, pH 7.2) at
193 65˚C with a radiolabeled Rep probe (primers: fw 5'-AACTGGACCAATGAGAACTTTCC-
194 3' and rv 5'-A AAAAGTCTTTGACTTCCTGCTT-3) or an ampicillin probe (primers: fw 5'-
195 AATCAGTGAGGCACCTATCTCAGC-3' and rv 5'- AACTCGGTCGCCGCATACACTATT-3') to
196 control for *Dpn*I digestion. Probes were labelled with the Prime-It RmT random
197 primer labeling kit from Stratagene and [^{32}P]dCTPs (PerkinElmer). The membranes
198 were exposed to a phosphorimager screen overnight. Images were acquired using a
199 Typhoon PhosphorImager (Molecular Dynamics) and analyzed with the ImageQuant
200 TL software (GE Healthcare Life Sciences).

201 **Fluorescence anisotropy DNA binding assay.**

202 Binding assays were performed using a fluorescein labelled 41-mer containing *AAVS1*
203 or p5 RBS sequences. The sequences used are 5'-
204 TGGCGGCGGTTGGGGCTCGGC*GCTCGCTCGCTC*GCTGGGCG-3' (*AAVS1*) and 5'-
205 ACCGGGCAAAATGGAGACCCTGCGTGCTCACTCGGGCTTAA-3' (p5) (40, 41). Rep68 WT
206 and mutant proteins at concentrations ranging from 5nM to 3µM were mixed with
207 DNA (5nM) in a final volume of 300 µl using the following buffer: 25mM HEPES (pH
208 7.0), 100mM NaCl, 1mM TECEP. Fluorescence readings were taken on a PC1
209 fluorimeter (ISS, Inc.) with excitation and emission filters at 490 and 520nm. Tubes
210 were equilibrated at 20°C for 20 minutes before measurement. Each anisotropy
211 point is the average of 10 measurements. Anisotropy is calculated as the ratio of the

difference between vertical and horizontal emission intensities over the total normalized intensity. The fraction of DNA bound (B) was calculated using the following equation:

$B = ([A]_x - [A]_{DNA})/([A]_{FINAL} - [A]_{DNA})$

Where $[A]_x$ represents the anisotropy measured at protein concentration X, $[A]_{DNA}$ is the anisotropy of free fluorescence DNA and $[A]_{FINAL}$ is the anisotropy at saturation. The data were fit to a single binding site model with Hill coefficient using the program Origin™ (Origin Labs). Each experiment was done in triplicate.

Helicase Assay

The substrate used in this assay is a heteroduplex DNA consisting of an 18-bp duplex region with a 10-nucleotide 3' tail at the bottom strand. The top strand (trap-DNA) is labelled at the 5' end with fluorescein, and is released upon unwinding. The sequences used are 5'-F-CATATGGAGCAGAACAGA-3' for the trap DNA and 5'-AGACAAGACGAGGTATACAAAAAAAAAA-3' for the complementary strand.

All reactions were performed in a buffer containing 25mM HEPES, 50mM NaCl (pH 7.0) at a total volume of 50µl. 1mM of protein was mixed with 0.5mM of double stranded F-DNA (18ADT10A) and 2.5µM of single stranded DNA (18s), and then added to the mix of buffer described above containing 5mM of both ATP and MgCl2. Reaction was incubated at 25˚C for 1 hr. EDTA was used to stop the reaction at a final concentration of 20µM. Aliquots of 10µl were loaded in a 12% bis-acrylamide gel (30%) (19:1) using 6X-loading dye (0.25 xylene cyanol FF, 30% glycerol). For the densitometry and analysis of the bands, a Gel Doc EZ Imager was used, together with

the automatic lane and band detection tool. Lane background subtraction, white illumination and an activation time of 300 sec were used for the analysis.

Supercoiled (sc) DNA nicking assay.

ScDNA nicking activity for Rep68 was assayed as described previously (14). Briefly, assays were performed in 30µl reactions containing 30mM Hepes-KOH (pH 7.5), 7mM MgCl2, 0.5mM DTT, 4mM ATP, 40mM creatine phosphate (Sigma), 1µg creatine phosphokinase (Sigma) in 15mM NaCl final concentration. 100ng scDNA plasmid and 200ng of purified His-Rep68 (or mutants) were added to the reactions. All samples were incubated for 1h at 37˚C; the reaction was terminated by adding 10µl of stop reaction (proteinase K [1.2µg/µl], 0.5% SDS and 30mM EDTA pH7.5) and incubating for 1 h at 37˚C. Samples were resolved in a 1% agarose gel (1X TAE), which was subsequently stained with ethidium bromide (0.3µg/ml) in 1X TAE. The scDNA plasmids used in this assay were pRVK (contains *AAVS1* nucleotide 1 to 3536bp) and a mutated version containing a mutant *trs* sequence (42).

Western Blot

Proteins were extracted from cells transfected and infected as described for the replication assays. Cells were lysed in RIPA buffer and the cleared lysate was run on a 12% acrylamide gel. The proteins were transferred onto a nitrocellulose membrane (GE healthcare) and immunoblotted using anti-Rep antibody (clone 303.9, Progen, 1/100 dilution), anti-Cap antibody (clone B1, ARP, 1/500) and anti-HSP90 antibody (polyclonal, Santa Cruz, 1/5000). All antibodies were incubated in blocking buffer (5% non fat dried milk in PBS containing 0.1% Tween 20). Images were acquired and analysed using an Imagequant apparatus (GE healthcare).

Real-time qRT-PCR

293T cells were transfected and infected as described for the replication assays. Total RNA was extracted using the RNeasy kit (Qiagen) after DNAseI (Qiagen) treatment for 15 minutes at 37˚C. Reverse transcription was performed using the High Capacity Reverse Transcription kit (Applied Biosystems). cDNA was quantified by real-time qPCR on an ABI PRISM system (Applied Biosystems) using the TaqMan Universal PCR master mix (Life Technologies and custom designed primer-probe mixes (Eurofins). Primers: p5 fw (5'-AACAAGGTGGTGGATGAGT-3'), p5 rv (5'-CGTTTACGCTCCGTGAGATT-3'), p19 fw (5'-TCACCAAGCAGGAAGTCAAAG-3'), p19 rv (5'-CCCGTTTGGGCTCACTTATATC-3'), p40 fw (5'-GGAAGCAAGGCTCAGAGAAA-3') and p40 rv (5'-CCTCTCTGGAGGTTGGTAGATA-3'). Probes: p5 (5'-FAM-ACGTGGTTGAGGTGGAGCATGAAT-TAM-3'), p19 (5'-FAM-ACGTGGTTGAGGTGGAGCATGAA-TAM-3') and p40 (5'-FAM-AGGAAATCAGGACAACCAATCCCGT-TAM-3'). Relative expression levels were determined with the ΔΔCt quantification method (43), using 18S ribosomal RNA (Taqman pre-developed assay reagents, human 18S rRNA, Applied Biosystems) as a housekeeping reference gene.

Results

Y224 forms hydrophobic interactions necessary for Rep68 oligomerization

Previous studies showed that Rep68 exists as a mixture of oligomers in solution. More specifically, two major populations have been observed by sedimentation velocity experiments, including a monomer-dimer peak that sediments at ~3S and oligomeric rings that sediment at 13S (25). We also showed that replacing the

281 tyrosine positioned at the C-terminal end of the linker in Rep68 by the smaller
282 residue alanine, disrupt its oligomerization; this presumably because of a reduction
283 in the surface exposed area (Figure 1A) (28). To further confirm this hypothesis we
284 mutated the tyrosine to phenylalanine (Phe), proline (Pro) or aspartic acid (Asp) and
285 performed sedimentation velocity experiments to study how these mutations affect
286 oligomerization. Figure 1B shows that replacement of the tyrosine with the small size
287 side-chain amino acids Pro and Asp has a drastic effect on the sedimentation profile
288 of Rep68. The 13S peak disappears, and the most prominent population present in
289 solution has a sedimentation coefficient around 5S, suggestive of smaller molecular
290 weight oligomers (Figure 1B). Exchanging the tyrosine with the bulky aromatic Phe
291 results in the appearance of two peaks, one with a sedimentation coefficient of 5S
292 similar to what we observed for the other mutants, and a second peak around 12S
293 that is indicative of the formation of larger oligomers. This 12S population, however,
294 has a smaller sedimentation coefficient than what we observe with Rep68,
295 potentially suggesting that the Y224F forms different oligomeric species. Figure 1C
296 shows the quantification of the 13S population that is formed in the presence of the
297 different mutations. Taken together, these results demonstrate that the bulky
298 aromatic character of the residue Y224 is pivotal for Rep68 oligomerization and
299 suggest that Y224 may participate in hydrophobic interactions as part of an
300 oligomeric interface.
301

Generation of a Rep68 oligomeric interface model

To further determine whether Y224 participates directly in forming an oligomeric interface, we modelled an oligomeric Rep dimer using the available structure of Rep40 (pdbid 1S9H), which spans residues 225-490 of Rep68 (33). We added the interdomain linker residues 217-224 to the known Rep40 structure as an extended α-helix based on secondary structure predictions (28), resulting in a Rep molecule containing residues 217 to 490 (Figure 2A). Two of the three molecules found in the asymmetric unit of Rep40 crystals form a pseudo-dimer. The interface formed in this dimer is similar to the oligomeric interface described for other SF3 helicase structures, but is not optimal to perform catalysis (44). We used this 'dimer', with the addition of the linker residues as our initial interface model, and we refined it by carrying out rigid body and side-chain conformation optimization using RosettaDock (45, 46). Strikingly, the top 10 models generated had almost identical interfaces as analysed by the program PISA (47), suggesting that our model was robust. A figure of the Rep interface model is shown in figure 2B. The interface buries a total of 1992 Å of solvent accessible area and includes residues from all the helices in the oligomerization domain (OD), the pre-sensor 1 β-hairpin (PS1βH), the β2β3 loop and residues from Walker A and Walker B motifs (Figure 2B). A closer analysis reveals that the modelled linker residues participate in the interface. In particular, the conserved aromatic residue Y224, which is at the end of the linker region, is an important component of the oligomeric interface. In agreement with the results shown in Figure 1, it participates in the formation of a hydrophobic pocket. Among the residues from the neighbouring subunit interacting with Y224, only I251 forms a

hydrophobic interaction whereas N254 contributes to a hydrogen bond via its main chain carbonyl oxygen (Figure 2C).

Mutations leading to disruption of the interface affect oligomerization of Rep68

Based on these observations, we generated a Rep68-I251A mutant and we assessed the consequences of mutating this residue alone or in combination with Y224 on Rep68 oligomerization. Figure 2D shows sedimentation velocity profiles illustrating that these mutants are mostly monomeric, thus validating the prediction from our model. Furthermore, the oligomerization defects, which we observed when testing the different Y224 mutants (Figure 1C) could also be explained by our modelled interface: smaller hydrophobic residues increase the solvent-accessible area and destabilize the interface, while a larger bulky residue maintains the hydrophobic pocket and possibly only affects the formation of the hydrogen bond between Y224 and N254, thus causing a milder defect. Finally, the Y224P mutation has the strongest effect, as it probably disrupts the helical character of this region and may affect the overall OD structure. To investigate whether the interface we identified is biologically relevant and thus involved in Rep functions, we assessed how the observed disruptions in the oligomerization profile of Rep68 affect the AAV life cycle.

Rep68 oligomerization is necessary to support the AAV life cycle

First, we verified if the aforementioned mutant Rep68 proteins are stable and localize correctly to the nucleus, where they support AAV replication. We transfected 293T cells with constructs expressing Rep68 or the oligomerization

mutants under the control of the CMV promoter, and assessed Rep68 protein stability and localization 48h post-transfection. Figure 3 demonstrates that all the mutants were expressed at levels comparable to those observed for wild-type (WT) Rep68 and translocated to the nucleus as expected.

Next, we assessed how disruptions in the oligomerization profile of Rep68 affect the AAV life cycle. We first compared the ability of the Rep oligomerization mutants to produce infectious AAV particles to that of WT Rep68 and Rep68*(Figure 4A). Rep68* or Rep* for short refers to the Cys151Ser mutant, which is functionally equivalent to WT Rep68 but prevents protein aggregation in solution (25). All mutant proteins have been generated in the context of Rep68*. As a negative control, we used the NTP-binding mutant K340H, which is deficient in ATPase and helicase activity and does not support AAV replication (48, 49). The K340H mutant, however, is still able to oligomerize and has been shown to have a dominant-negative phenotype (48-50). Recombinant AAV2-GFP particles were produced in 293T cells by transfection of plasmids encoding the adenovirus helper functions, AAV2 Cap, and WT Rep68, Rep68* or the interface mutants. Increasing volumes of supernatant collected from the cultures of AAV-producing cells were added to HeLa cells in order to assess the infectivity of the produced virus. Figure 4A shows that the Y224A Rep mutant does not support the production of infectious AAV particles as was previously reported by us (28). The more conservative mutation, Y224F, which retains the potential to partially oligomerize (Figure 1C), is severely impaired but is not entirely deficient in producing infectious AAV. Mutating I251 to alanine on the opposite side of the predicted interface, however, replicates the phenotype

observed with the Y224A mutant. Not surprisingly, the double mutant Y224A-I251A also fails to produce infectious AAV particles (Figure 4A).

To evaluate if the failure to produce infectious particles was due to a defect in AAV DNA replication, we determined the number of AAV genomes in the 293T producer cells by qPCR (Figure 4B) and studied the replicative intermediates formed during AAV replication by Southern blot (Figure 4C). Both assays confirmed that the interface mutants Y224A, I251A and the double mutant Y224A-I251A all fail to support AAV DNA replication. Similarly to what we observed in the infectious particle production assay, the Y224F mutant supports AAV replication, but at significantly lower levels than observed for Rep WT or Rep*. In addition, Figure 4C shows that replication in the presence of the Y224F mutant results in the formation of the expected replicative intermediates. Background replication can be observed in the absence of adenovirus due to the presence of E1A and E1B in 293T cells (51). Altogether, these results suggest that the oligomerization interface mutants fail to sustain AAV DNA replication and therefore cannot support the production of infectious AAV particles.

Rep68 oligomerization mutants are deficient in RBS-specific DNA binding and site- and strand-specific nicking but maintain the ability to unwind unspecific DNA substrates

In order to determine the cause of the replication defect of the oligomerization-deficient mutants, we assessed various biochemical activities *in vitro*. Rep has three well-characterised enzymatic functions – RBS-specific DNA binding, *trs* nicking and

ATP-dependent DNA unwinding – all of which are necessary for AAV DNA replication and targeted genome integration. Table 1 shows the binding constants of Rep68* and mutant Rep68* proteins on p5 and *AAVS1* RBS-containing double stranded DNA substrates. As expected, Rep68* and the control Rep68*-K340H both efficiently bind the specific DNA substrates (25, 52). The mutant that has retained some residual replication potential, Rep68*-Y224F, also efficiently binds the p5 and *AAVS1* DNA substrates. All other mutants, however, have lost the ability to bind both DNA substrates, with the exception of Rep68*-I251 that has maintained its ability to bind *AAVS1*-containing DNA, albeit with a 10-fold lower affinity than its WT counterpart. These results suggest that some level of oligomerization is necessary for efficient RBS-specific DNA binding by Rep68, and that the oligomeric properties of the mutant Rep68*-Y224F are sufficient for DNA binding.

To test the ability of the Rep68 mutants to unwind non-specific DNA, we performed a fluorescence-based helicase assay. Somewhat surprisingly, all mutants but the control K340H mutant exhibit similar helicase activity on a heteroduplex non-specific DNA substrate (Figure 5). These results suggest that under these experimental conditions Rep68 can unwind DNA even in the absence of large complexes, or alternatively that an oligomeric complex that is stabilized by a different interface is necessary for Rep-mediated DNA unwinding. Strand- and site-specific nicking activity, however, appears to diminish strongly when oligomerization is disrupted in Rep68. As shown in Figures 6A and 6B, the oligomerization-deficient Rep mutants that fail to bind specific DNA also fail to nick supercoiled plasmid DNA containing RBS and *trs* sequences. The Rep68*-Y224F mutant, despite retaining the ability to bind specific DNA substrates, only shows some residual nicking activity.

Rep oligomerization is important for transcriptional regulation of AAV genes.

In addition to their role in AAV DNA replication, the Rep proteins coordinate the temporal regulation of transcription of the viral genome during the AAV life cycle. In the absence of helper virus, the Rep proteins participate in repressing transcription from the three viral promoters p5, p19 and p40, ensuring minute levels of expression of the viral proteins. In the presence of helper virus, i.e. during a productive infection, repression of the p5 promoter is lifted by the adenoviral E1A protein (53) and binding of Rep to the p5 promoter or the ITRs leads to transactivation of the p19 and p40 promoters (54-56). The p5 promoter itself is also controlled by Rep, which can act both as a repressor or an activator through binding at the p5 or ITR RBS, respectively (56, 57). The net result is a self-regulatory loop that generates protein levels that are tightly controlled and are optimal for AAV replication and packaging (58). Two mechanisms of Rep-mediated repression have been identified: direct repression through binding at the RBS in the p5 promoter, and indirect repression that requires the ATPase activity of Rep (19). In light of the dependence of transcriptional regulation on Rep binding to the p5 promoter and ITR, we assessed whether the oligomerization mutants also displayed defects in transcriptional activity resulting in altered protein expression levels. As a control we used again the K340H ATPase mutant, which has been shown to lead to the expression of exceedingly high Rep protein levels under conditions permissive for AAV replication (48). Cells were transfected with the various AAV infectious plasmids, and Rep and Cap protein levels were determined in the presence and absence of adenovirus co-infection. In the presence of WT Rep and the absence of adenovirus, we expect low

levels of both Rep and Cap proteins. In the presence of adenovirus, Rep protein levels peak around 30h after infection and then slowly decrease, while Cap levels increase with viral DNA replication (56). Because we harvested the cells 72h post-transfection, we expect to see only slightly higher levels of the Rep proteins but significantly higher Cap expression when compared to cells that were not co-infected with adenovirus. As shown in Figure 7A, we observed strikingly high Rep protein levels in cells transfected with the mutated Rep proteins, including the control K340H mutant. Once more, we observed that the Y224F mutant showed an intermediate phenotype represented by a very modest increase in Rep protein expression levels (Figure 7A). The Cap protein levels, on the other hand, were found to be lower in cells expressing the mutant proteins: only in the presence of Rep proteins that support the AAV life cycle we detected high Cap levels, with the sole exception being the K340H mutant (Figure 7A). The same trend was observed both in the presence and the absence of adenovirus infection, although the Cap levels were significantly lower in the absence of helper virus. These results suggest that the oligomerization mutants fail to regulate the expression levels of the AAV proteins, most likely by failing to autoregulate the p5 promoter through RBS binding. In view of these important differences in protein amounts, we assessed the levels of AAV transcripts by RT-PCR (Figure 6C). Because all AAV RNAs use the same polyadenylation signal, we were not able to quantify the p19 and p40 transcripts separately from the p5 transcripts. As expected, with all primer sets used – targeting respectively p5, p5 and p19 and all major AAV mRNA – we observed an increase in mRNA levels in response to adenovirus co-infection in the presence of Rep proteins that support AAV replication. In the presence of the Y224F mutant, the response to

466 adenovirus is still present, but is nevertheless reduced when compared to WT Rep.
467 The oligomerization-deficient mutants Y224, I251, and Y224-I251, which are unable
468 to bind the RBS at the p5 promoter, have higher basal mRNA levels, varying between
469 2- and 10-fold, and do not respond to adenovirus infection. In the context of
470 adenovirus co-infection, however, the differences in mRNA levels do not correlate
471 with those observed for the protein levels, suggesting that changes in post-
472 transcriptional regulation are also contributing to the altered protein expression
473 levels. Rep-mediated post-transcriptional regulation has been observed before, but
474 its mechanism remains unknown (59). The K340H mutant, which oligomerizes but
475 fails to support AAV replication, has considerably higher basal mRNA levels as
476 compared to WT Rep, possibly explaining the very high Rep and Cap protein amounts
477 observed, and the presence of adenovirus does not lead to a clear change in mRNA
478 levels. Taken together, our data support a model in which Rep oligomerization is
479 important for the gene regulatory function of Rep, potentially through p5 RBS
480 binding, which is necessary to achieve an appropriate transcription profile.

481

482 **Discussion**

483 The limited genome capacity of small viruses such as adeno-associated virus has
484 driven the evolution of highly multifunctional non-structural proteins that combine
485 several enzymatic functions necessary to support the viral life cycle. In AAV, the Rep
486 proteins are responsible for orchestrating the entire viral life cycle, from
487 transcriptional regulation, to replication and packaging, as well as Rep-mediated
488 integration. The combination of several enzymatic functions, including DNA binding,

nicking und unwinding, and the ability to interact with a multitude of DNA substrates and proteins, allows the Rep proteins to support replication. However, the coordination of all these functions would require a tightly controlled system, which we envision could be provided by the different oligomeric states the protein assumes. During the AAV life cycle, Rep has to catalyse reactions on different DNA substrates, including initiation of DNA replication, recognition and nicking of the *trs*, and binding to the p5 promoter in order to provide transcriptional regulation. It has been shown that Rep can form different oligomeric species *in vitro*, both in the absence of DNA and in the presence of different DNA substrates (24, 25), allowing for an additional layer of regulation of the Rep activities during the AAV life cycle. To fully understand the mechanism of action of Rep on its different substrates, it is essential to identify the oligomeric complexes formed with the different DNA molecules. For example, it has been shown that Rep68 forms a double-octameric ring in the presence of ssDNA as well as on forked helicase substrates (24), whereas other reports have suggested that Rep68 forms hexamers when bound to dsDNA (50, 60). However, the importance of these complexes for the viral life cycle has not been formally addressed, and while it is clear that Rep oligomerization is functionally relevant, data on possible oligomerization interfaces remain scarce. Smith and colleagues identified two regions – residues 151-188 and 334-347 – that when deleted disrupt Rep oligomerization, however they did not investigate the functional consequences of these deletions (50). Intriguingly, residues 334-346 include the ATP binding site that has been shown to be part of the oligomeric interface in PV-E1 and SV40-LTag hexamers. A more recent report showed that one residue, R107, initially identified for its role in integration, origin binding and nicking, and shown to be in

direct contact with origin DNA, is also essential for oligomerization (24, 61). Finally, we and others have shown that the interdomain linker of Rep78/68, and in particular the Y224 residue, is critical for Rep oligomerization (28, 32).

Building on our previous studies, we further characterized the role of residue Y224 in Rep oligomerization. Substitution of Y224 using residues with different properties had varying consequences on Rep68* oligomerization. More specifically, small hydrophobic residue substitutions severely impaired oligomerization, while the more conservative Y224F mutant retained the ability to oligomerize (Figure 1), suggesting that Y224 participates in the formation and the stabilization of a hydrophobic interface. This hypothesis was supported by a model of a dimeric Rep-Rep interaction built from the pseudo-dimer observed in the crystal structure of Rep40 using an extended Rep40 molecule (33). In this model, a large interface is formed that resembles the interface formed by the PV-E1 protein and includes residues from the Walker A, Walker B, PS1βH and β2β3 (Figure 2). Furthermore, all the helices in the OD are also participating in the interface and form a hydrophobic pocket, emphasizing the relevance of this sub-domain in Rep oligomerization. More specifically, linker residue Y224 on the extended α-helix 1 of one Rep molecule is interacting with residue I251 and with the main chain carbonyl oxygen of residue N254 on α-helix 3 of the other Rep molecule (Figure 2C). Mutating I251 to alanine in Rep68, alone or in combination with Y224A, confirmed that this residue is important for Rep68 oligomerization. Importantly, because none of the residues located in the OD that we identified as participating in the oligomeric interface are part of the described catalytic sites within the Rep proteins, the consequences of these mutations on the functions of the Rep proteins are likely to be due to

oligomerization defects. We showed that oligomerization-deficient mutants Y224A, I251A and Y224A-I251A are unable to replicate AAV DNA and fail to support the production of recombinant AAV. Our data suggest that these defects are caused by the loss of DNA binding and origin nicking activities by the mutant Rep proteins, and confirm that Rep oligomerization is critical for its function in support of the AAV life cycle. We also assessed the consequences on Rep function of a more conservative mutation, Y224F. This substitution maintains the bulky aromatic character of the residue, a feature that is conserved in the OD of SF3 helicases and other related proteins (28). The Rep68*-Y224F mutant retains the ability to oligomerize, but forms the large 13S complexes less efficiently than observed with Rep68* (Figure 1). Interestingly, the binding of RBS-containing DNA does not appear to be compromised by this mutation; nicking activity on the other hand is severely impaired, possibly explaining the observed low levels of viral replication (Figure 6). In view of the oligomeric behavior of the Y224F mutant, these results suggest that this mutant retains the ability to form an oligomeric complex sufficient for RBS-mediated DNA binding, but fails to promote the subsequent DNA nicking step. How this transition is affected, however, is not clear. One intriguing possibility is that the initial Rep binding to the RBS and melting of the origin promotes the recruitment of further Rep78/68 molecules and the assembly of a second, larger, Rep-DNA complex that is necessary for the nicking reaction. Residue Y224, and more generally the OD, could help stabilize the formation of this complex, allowing a shift in the interaction with the origin DNA to allow the *trs* nicking reaction to take place. The Y224F mutation had previously been identified in a study by Walker et al. to be important for Rep function (62). In contrast to our findings, however, the authors reported that

561 the Y224F mutant was deficient in ITR binding, endonuclease, DNA helicase and
562 ATPase activities. The cause of this difference may be due to a different
563 experimental strategy, or could possibly be explained by the presence of a maltose-
564 binding protein (MBP) tag (62), which may affect the already weakened
565 oligomerization potential of Rep68-Y224F.
566 SF3 helicases are thought to function as oligomeric complexes, as is the case for PV-
567 E1 and SV40-LTag that form active hexameric complexes. Surprisingly, all the
568 oligomerization-deficient mutants described here were still able to unwind a
569 heteroduplex DNA substrate (Figure 5). This suggests that interaction with and the
570 unwinding of this DNA substrate does not require the formation of large Rep
571 oligomers, or alternatively that the presence of heteroduplex DNA and ATP stabilizes
572 the formation of an oligomeric complex independently from the oligomeric interface
573 described here. Rep40 and Rep52 are mostly monomeric and retain helicase activity,
574 albeit to a lower level than that observed with the large Rep proteins (63). In our
575 previous report we have introduced the possibility that AAV Rep proteins have
576 evolved two distinct helicase modes (28): one that parallels the helicase activity of
577 other SF3 helicases, requires oligomeric rings and is performed by the large Rep
578 proteins, and one that only requires a transient dimerization and is characteristic for
579 the activity of the small Rep isoforms. Thus, it is plausible that the mutants described
580 here are still able to unwind DNA through the same mechanism used by Rep52/40,
581 but they would not support the unwinding of a substrate that requires the helicase
582 activity from oligomeric rings. Based on the different functions in the AAV life cycle
583 of the large and small Rep proteins, it is tempting to suggest that melting of the AAV
584 origin, which is mediated by the large Rep proteins, requires the formation of a

stable oligomer that unwinds DNA by a mechanism analogous to that described for other SF3 helicases, while packaging AAV genomes into the viral capsids, which is efficiently carried out by the small Rep proteins, may proceed through a different helicase mode.

In addition to their role in supporting AAV DNA replication, the enzymatic activities of the Rep proteins are also essential for the correct transcriptional regulation of viral and cellular transcripts. Because it is known that the levels of Rep proteins are tightly regulated – and not simply maximized – to achieve efficient AAV DNA replication (58, 64), we assessed whether the expression levels of AAV proteins was affected by the oligomerization mutants. The K340H Rep mutant has been shown to fail in appropriately regulating the expression of the AAV genes, suggesting that the ATPase/helicase activity of Rep is involved in transcriptional regulation (48). A different mechanism of transcriptional repression, dependent on the RBS binding activity of Rep78/68 but not on its ATPase function, has also been demonstrated (20). In this study, we show that regulation of AAV gene expression is impaired in the presence of Rep oligomerization mutants that do not bind p5 or *AAVS1* DNA, and our results suggest that oligomerization of the large Rep proteins is necessary for the correct regulation of the transcription of all AAV promoters (figure 7). More specifically, the oligomerization-deficient mutants fail to induce transcription of the viral promoters upon infection with the helper virus adenovirus and, in addition, both large and small Rep protein levels increase substantially in the presence of oligomerization mutants. The presence of the Y224A mutation creates a stronger Kozak sequence at the p19 promoter as compared to WT Rep68, but this is not sufficient to explain the differences in protein levels observed, in particular for

609 Rep78. In addition, the increase in Rep52 expression in the presence of the Y224A-
610 I251A mutations is modest compared to that observed for the Y224A alone. The
611 differences in protein levels observed in presence of adenovirus, however, cannot be
612 explained by the RNA levels alone, suggesting that there is some level of post-
613 transcriptional control that may also be Rep-dependent. A function for AAV Rep in
614 this context was previously suggested in a study by Trempe and Carter, where it was
615 observed that regulation of gene expression at a transcriptional level alone was not
616 sufficient to explain differences in protein levels (59). Our data also support a role for
617 an oligomeric complex of Rep in regulating protein levels post-transcriptionally.
618 Understanding the mechanism behind this potential uncharacterized function of AAV
619 Rep proteins may reveal a new layer of complexity in the role that the Rep proteins
620 play in coordinating the AAV life cycle.
621 In conclusion, our study identifies and describes an essential Rep-Rep protein
622 interface that is involved in the formation of Rep complexes and demonstrates its
623 functional relevance throughout the AAV life cycle. Our study focuses on residues
624 that are part of the α-helical bundle located upstream of the helicase domain, and
625 strengthen the role of this sub-domain of Rep as a *bona fide* oligomerization domain.
626 The identification of the oligomeric interfaces of AAV Rep like the one described
627 here, and further structural and functional characterisation of Rep oligomeric
628 complexes, particularly in the presence of different DNA substrates, will provide
629 additional insights into the molecular mechanisms of Rep-mediated transcriptional
630 regulation and AAV DNA replication, as well as Rep-mediated integration.
631

References

1. **Weitzman MD, Linden RM.** 2011. Adeno-associated virus biology. Methods in molecular biology **807:**1-23.

2. **Green MR, Roeder RG.** 1980. Transcripts of the adeno-associated virus genome: mapping of the major RNAs. J Virol **36:**79-92.

3. **Srivastava A, Lusby EW, Berns KI.** 1983. Nucleotide sequence and organization of the adeno-associated virus 2 genome. J Virol **45:**555-564.

4. **Smith RH, Kotin RM.** 1998. The Rep52 gene product of adeno-associated virus is a DNA helicase with 3'-to-5' polarity. J Virol **72:**4874-4881.

5. **Im DS, Muzyczka N.** 1990. The AAV origin binding protein Rep68 is an ATP-dependent site-specific endonuclease with DNA helicase activity. Cell **61:**447-457.

6. **Owens RA, Weitzman MD, Kyostio SR, Carter BJ.** 1993. Identification of a DNA-binding domain in the amino terminus of adeno-associated virus Rep proteins. J Virol **67:**997-1005.

7. **Davis MD, Wu J, Owens RA.** 2000. Mutational analysis of adeno-associated virus type 2 Rep68 protein endonuclease activity on partially single-stranded substrates. J Virol **74:**2936-2942.

8. **Di Pasquale G, Stacey SN.** 1998. Adeno-associated virus Rep78 protein interacts with protein kinase A and its homolog PRKX and inhibits CREB-dependent transcriptional activation. J Virol **72:**7916-7925.

9. **Berthet C, Raj K, Saudan P, Beard P.** 2005. How adeno-associated virus Rep78 protein arrests cells completely in S phase. Proc Natl Acad Sci U S A **102:**13634-13639.

10. **King JA, Dubielzig R, Grimm D, Kleinschmidt JA.** 2001. DNA helicase-mediated packaging of adeno-associated virus type 2 genomes into preformed capsids. EMBO J **20:**3282-3291.

11. **Dutheil N, Linden RM.** 2006. Site-specific integration by adeno-associated virus, p 214-236. *In* Kerr JR, Cotmore SF, Bloom ME, Linden RM, Parrish CR (ed), Parvoviruses. Hodder Arnold, London.

12. **Ward P.** 2006. Replication of adeno-associated virus DNA, p 189-211. *In* Kerr JR, Cotmore SF, Bloom ME, Linden RM, Parrish CR (ed), Parvoviruses. Hodder Arnold, London.

13. **Brister JR, Muzyczka N.** 2000. Mechanism of Rep-mediated adeno-associated virus origin nicking. J Virol **74:**7762-7771.

14. **Petri K, Gabriel R, Agundez L, Fronza R, Afzal S, Kaeppel C, Linden RM, Henckaerts E, Schmidt M.** 2015. Presence of a trs-Like Motif Promotes Rep-Mediated Wild-Type Adeno-Associated Virus Type 2 Integration. J Virol **89:**7428-7432.

15. **Weitzman MD, Kyostio SR, Kotin RM, Owens RA.** 1994. Adeno-associated virus (AAV) Rep proteins mediate complex formation between AAV DNA and its integration site in human DNA. Proc Natl Acad Sci U S A **91:**5808-5812.

16. **Urcelay E, Ward P, Wiener SM, Safer B, Kotin RM.** 1995. Asymmetric replication in vitro from a human sequence element is dependent on adeno-associated virus Rep protein. J Virol **69:**2038-2046.

17. **Brister JR, Muzyczka N.** 1999. Rep-mediated nicking of the adeno-associated virus origin requires two biochemical activities, DNA helicase activity and transesterification. J Virol **73:**9325-9336.

18. **Horer M, Weger S, Butz K, Hoppe-Seyler F, Geisen C, Kleinschmidt JA.** 1995. Mutational analysis of adeno-associated virus Rep protein-mediated inhibition of heterologous and homologous promoters. J Virol **69:**5485-5496.

19. **Kyostio SR, Wonderling RS, Owens RA.** 1995. Negative regulation of the adeno-associated virus (AAV) P5 promoter involves both the P5 rep binding site and the consensus ATP-binding motif of the AAV Rep68 protein. J Virol **69:**6787-6796.

20. **Dutheil N, Smith SC, Agundez L, Vincent-Mistiaen ZI, Escalante CR, Linden RM, Henckaerts E.** 2014. Adeno-associated virus Rep represses the human integration site promoter by two pathways that are similar to those required for the regulation of the viral p5 promoter. J Virol **88:**8227-8241.

21. **Hickman AB, Dyda F.** 2005. Binding and unwinding: SF3 viral helicases. Curr Opin Struct Biol **15:**77-85.

22. **Chandler M, de la Cruz F, Dyda F, Hickman AB, Moncalian G, Ton-Hoang B.** 2013. Breaking and joining single-stranded DNA: the HUH endonuclease superfamily. Nat Rev Microbiol **11:**525-538.

23. **Gonzalez-Prieto C, Agundez L, Linden RM, Llosa M.** 2013. HUH site-specific recombinases for targeted modification of the human genome. Trends Biotechnol **31:**305-312.

24. **Mansilla-Soto J, Yoon-Robarts M, Rice WJ, Arya S, Escalante CR, Linden RM.** 2009. DNA structure modulates the oligomerization properties of the AAV initiator protein Rep68. PLoS Pathog **5:**e1000513.

25. **Zarate-Perez F, Mansilla-Soto J, Bardelli M, Burgner JW, 2nd, Villamil-Jarauta M, Kekilli D, Samso M, Linden RM, Escalante CR.** 2013. Oligomeric properties of adeno-associated virus Rep68 reflect its multifunctionality. J Virol **87:**1232-1241.

26. **Sedman J, Stenlund A.** 1998. The papillomavirus E1 protein forms a DNA-dependent hexameric complex with ATPase and DNA helicase activities. J Virol **72:**6893-6897.

27. **Li D, Zhao R, Lilyestrom W, Gai D, Zhang R, DeCaprio JA, Fanning E, Jochimiak A, Szakonyi G, Chen XS.** 2003. Structure of the replicative helicase of the oncoprotein SV40 large tumour antigen. Nature **423:**512-518.

28. **Zarate-Perez F, Bardelli M, Burgner JW, 2nd, Villamil-Jarauta M, Das K, Kekilli D, Mansilla-Soto J, Linden RM, Escalante CR.** 2012. The interdomain linker of AAV-2 Rep68 is an integral part of its oligomerization domain: role of a conserved SF3 helicase residue in oligomerization. PLoS Pathog **8:**e1002764.

29. **Titolo S, Pelletier A, Pulichino AM, Brault K, Wardrop E, White PW, Cordingley MG, Archambault J.** 2000. Identification of domains of the human papillomavirus type 11 E1 helicase involved in oligomerization and binding to the viral origin. J Virol **74:**7349-7361.

30. **Loeber G, Stenger JE, Ray S, Parsons RE, Anderson ME, Tegtmeyer P.** 1991. The zinc finger region of simian virus 40 large T antigen is needed for hexamer assembly and origin melting. Journal of virology **65:**3167-3174.

31. **Boer DR, Ruiz-Maso JA, Lopez-Blanco JR, Blanco AG, Vives-Llacer M, Chacon P, Uson I, Gomis-Ruth FX, Espinosa M, Llorca O, del Solar G, Coll M.** 2009. Plasmid replication initiator RepB forms a hexamer reminiscent of ring helicases and has mobile nuclease domains. EMBO J **28:**1666-1678.

32. **Maggin JE, James JA, Chappie JS, Dyda F, Hickman AB.** 2012. The amino acid linker between the endonuclease and helicase domains of adeno-associated virus type 5 Rep plays a critical role in DNA-dependent oligomerization. J Virol **86:**3337-3346.

33. **James JA, Escalante CR, Yoon-Robarts M, Edwards TA, Linden RM, Aggarwal AK.** 2003. Crystal structure of the SF3 helicase from adeno-associated virus type 2. Structure **11:**1025-1035.

34. **Schuck P.** 2003. On the analysis of protein self-association by sedimentation velocity analytical ultracentrifugation. Anal Biochem **320:**104-124.

35. **Vistica J, Dam J, Balbo A, Yikilmaz E, Mariuzza RA, Rouault TA, Schuck P.** 2004. Sedimentation equilibrium analysis of protein interactions with global implicit mass conservation constraints and systematic noise decomposition. Anal Biochem **326:**234-256.

36. **Zolotukhin S, Potter M, Hauswirth WW, Guy J, Muzyczka N.** 1996. A "humanized" green fluorescent protein cDNA adapted for high-level expression in mammalian cells. J Virol **70:**4646-4654.

37. **Streck CJ, Dickson PV, Ng CY, Zhou J, Hall MM, Gray JT, Nathwani AC, Davidoff AM.** 2006. Antitumor efficacy of AAV-mediated systemic delivery of interferon-beta. Cancer Gene Ther **13:**99-106.

38. **Laughlin CA, Tratschin JD, Coon H, Carter BJ.** 1983. Cloning of infectious adeno-associated virus genomes in bacterial plasmids. Gene **23:**65-73.

39. **Hirt B.** 1969. Replicating molecules of polyoma virus DNA. Journal of molecular biology **40:**141-144.

40. **McCarty DM, Pereira DJ, Zolotukhin I, Zhou X, Ryan JH, Muzyczka N.** 1994. Identification of linear DNA sequences that specifically bind the adeno-associated virus Rep protein. J Virol **68:**4988-4997.

41. **Linden RM, Winocour E, Berns KI.** 1996. The recombination signals for adeno-associated virus site-specific integration. Proc Natl Acad Sci U S A **93:**7966-7972.

42. **Lamartina S, Ciliberto G, Toniatti C.** 2000. Selective cleavage of AAVS1 substrates by the adeno-associated virus type 2 rep68 protein is dependent on topological and sequence constraints. J Virol **74:**8831-8842.

43. **Schmittgen TD, Livak KJ.** 2008. Analyzing real-time PCR data by the comparative C(T) method. Nature protocols **3:**1101-1108.

44. **James JA, Aggarwal AK, Linden RM, Escalante CR.** 2004. Structure of adeno-associated virus type 2 Rep40-ADP complex: insight into nucleotide recognition and catalysis by superfamily 3 helicases. Proc Natl Acad Sci U S A **101:**12455-12460.

45. **Gray JJ, Moughon S, Wang C, Schueler-Furman O, Kuhlman B, Rohl CA, Baker D.** 2003. Protein-protein docking with simultaneous optimization of rigid-body displacement and side-chain conformations. J Mol Biol **331:**281-299.

46. **Lyskov S, Gray JJ.** 2008. The RosettaDock server for local protein-protein docking. Nucleic Acids Res **36:**W233-238.

47. **Krissinel E, Henrick K.** 2007. Inference of macromolecular assemblies from crystalline state. J Mol Biol **372:**774-797.

48. **Chejanovsky N, Carter BJ.** 1990. Mutation of a consensus purine nucleotide binding site in the adeno-associated virus rep gene generates a dominant negative phenotype for DNA replication. J Virol **64:**1764-1770.

49. **Kyostio SR, Owens RA.** 1996. Identification of mutant adeno-associated virus Rep proteins which are dominant-negative for DNA helicase activity. Biochem Biophys Res Commun **220:**294-299.

50. **Smith RH, Spano AJ, Kotin RM.** 1997. The Rep78 gene product of adeno-associated virus (AAV) self-associates to form a hexameric complex in the presence of AAV ori sequences. J Virol **71:**4461-4471.

51. **Wang XS, Srivastava A.** 1998. Rescue and autonomous replication of adeno-associated virus type 2 genomes containing Rep-binding site mutations in the viral p5 promoter. Journal of virology **72:**4811-4818.

52. **Owens RA, Trempe JP, Chejanovsky N, Carter BJ.** 1991. Adeno-associated virus rep proteins produced in insect and mammalian expression systems: wild-type and dominant-negative mutant proteins bind to the viral replication origin. Virology **184:**14-22.

53. **Shi Y, Seto E, Chang LS, Shenk T.** 1991. Transcriptional repression by YY1, a human GLI-Kruppel-related protein, and relief of repression by adenovirus E1A protein. Cell **67:**377-388.

54. **Pereira DJ, Muzyczka N.** 1997. The adeno-associated virus type 2 p40 promoter requires a proximal Sp1 interaction and a p19 CArG-like element to facilitate Rep transactivation. J Virol **71:**4300-4309.

55. **Pereira DJ, Muzyczka N.** 1997. The cellular transcription factor SP1 and an unknown cellular protein are required to mediate Rep protein activation of the adeno-associated virus p19 promoter. J Virol **71:**1747-1756.

56. **Weger S, Wistuba A, Grimm D, Kleinschmidt JA.** 1997. Control of adeno-associated virus type 2 cap gene expression: relative influence of helper virus, terminal repeats, and Rep proteins. J Virol **71:**8437-8447.

57. **Labow MA, Hermonat PL, Berns KI.** 1986. Positive and negative autoregulation of the adeno-associated virus type 2 genome. J Virol **60:**251-258.

58. **Li J, Samulski RJ, Xiao X.** 1997. Role for highly regulated rep gene expression in adeno-associated virus vector production. J Virol **71:**5236-5243.

59. **Trempe JP, Carter BJ.** 1988. Regulation of adeno-associated virus gene expression in 293 cells: control of mRNA abundance and translation. J Virol **62:**68-74.

60. **Dignam SS, Correia JJ, Nada SE, Trempe JP, Dignam JD.** 2007. Activation of the ATPase activity of adeno-associated virus Rep68 and Rep78. Biochemistry **46:**6364-6374.

61. **Urabe M, Hasumi Y, Kume A, Surosky RT, Kurtzman GJ, Tobita K, Ozawa K.** 1999. Charged-to-alanine scanning mutagenesis of the N-terminal half of adeno-associated virus type 2 Rep78 protein. J Virol **73:**2682-2693.

62. **Walker SL, Wonderling RS, Owens RA.** 1997. Mutational analysis of the adeno-associated virus Rep68 protein: identification of critical residues necessary for site-specific endonuclease activity. J Virol **71:**2722-2730.

63. **Collaco RF, Kalman-Maltese V, Smith AD, Dignam JD, Trempe JP.** 2003. A biochemical characterization of the adeno-associated virus Rep40 helicase. J Biol Chem **278:**34011-34017.

64. **Pereira DJ, McCarty DM, Muzyczka N.** 1997. The adeno-associated virus (AAV) Rep protein acts as both a repressor and an activator to regulate AAV transcription during a productive infection. J Virol **71:**1079-1088.

805

806 **Figure Legends**

807

808 **Figure 1: Role of Tyr224 in Rep68 oligomerization. (A)** Schematic diagram of Rep68,
809 with the OBD in green and the helicase domain (HD) in blue. The sequence of the
810 linker is shown and Y224 is highlighted in red. **(B)** Sedimentation velocity analysis of
811 Rep68 and different Y224 mutants. **(C)** Quantification of the amount of 13S species
812 formed by Y224 mutants.

813

814 **Figure 2: Model of a Rep-Rep interface. (A)** Ribbon representation of Rep40
815 extended to residue 217 as an α-helix highlighted in dark blue. The Walker A, Walker
816 B, PS1βH, the β2β3 loop and OD are indicated. **(B)** Model of a dimeric Rep complex.
817 The structures participating in the interface are highlighted in red and magenta. **(C)**
818 Close-up of the interactions formed by residue Y224, including the hydrophobic
819 interaction with I251 and the hydrogen bond with the backbone of N254. **(D)**
820 Sedimentation velocity analysis of Rep68 mutants I251A and Y224A-I251A (YI
821 mutant).

822

823 **Figure 3: The interface mutants are stable and localize to the nucleus. (A)** Rep68
824 WT or mutants were transfected in 293T cells and tested for expression levels by
825 Western blotting 36h post-transfection. **(B)** 293T cells transfected with Rep68 WT or
826 mutants were assessed 48h post-transfection by immunostaining.

827

828

Figure 4: Interface mutants do not support the AAV life cycle. (A) Increasing volumes of supernatant from 293T cells producing recombinant AAV-GFP in the presence of Rep WT or mutants were used to infect HeLa cells, and the percentage of GFP-positive cells was determined by FACS analysis. Data from three independent experiments represented as mean ± SEM. **(B)** AAV DNA replication under permissive (+ Ad) and non-permissive (no Ad) conditions was quantified by quantitative PCR. Data from three experiments were normalized to the Rep*+Ad condition, and are represented as mean ± SEM. **(C)** AAV replicative intermediates generated in the same conditions as in (B) were visualized by Southern blotting using a Rep-specific probe. RfM: monomeric replicative form. RfD: dimeric replicative form.

Figure 5: Comparison of the helicase activity of the interface mutants. The ability of the Rep68 interface mutants to unwind a fluorescein labeled heteroduplex DNA substrate was assayed. Data from three experiments represented as mean ± SEM.

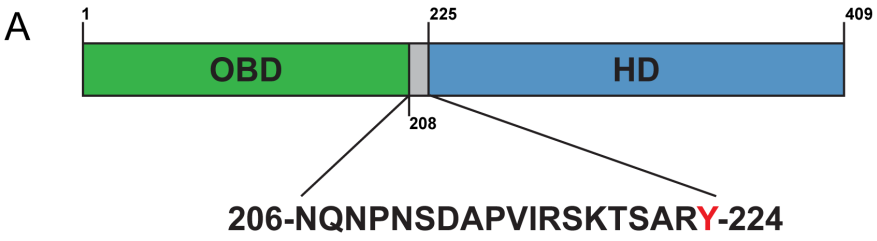
Figure 6: Rep-mediated nicking of supercoiled plasmid. Supercoiled (SC) plasmid DNA containing a RBS and a *trs* was mixed with Rep68* or the interface mutants. If the endonuclease activity is intact Rep will nick and relax the plasmid conformation to an open circular (OC) form, which can readily be distinguished by agarose gel electrophoresis. A *trs* mutant (*TRSmut*) plasmid that is not nicked by Rep was used as control. Untreated indicates DNA that was left untouched, while mock treated indicates plasmid that was incubated in reaction buffer 1h at 37˚C in the absence of protein. **(A)** Representative agarose gel electrophoresis image. **(B)** Quantification of four independent nicking experiments. Data represented as mean ± SEM.

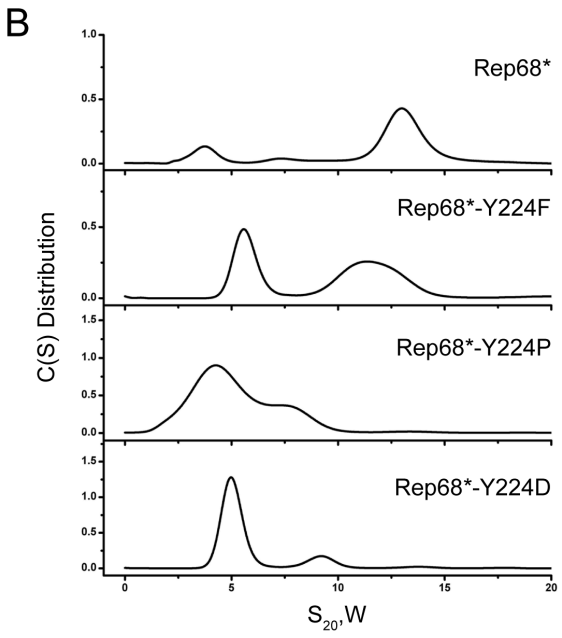
Figure 7: Rep oligomerization is important for transcriptional regulation of AAV genes. (A) Western blot showing Rep and Cap protein levels under conditions permissive (+ adenovirus) and non permissive for AAV replication. The first lane of the left panel is equivalent to the third lane in the right panel. **(B)** Transcription levels of AAV genes were analyzed under the same conditions as in (A). RNA levels were measured by RT-PCR using three primer-probe mixes detecting RNA from the p5 promoter only, from p5 and p19, and from p5, p19 and p40. The fold change was calculated relative to the mRNA levels in the presence of WT Rep but in the absence of adenovirus. Data from three experiments represented as mean ± SEM.

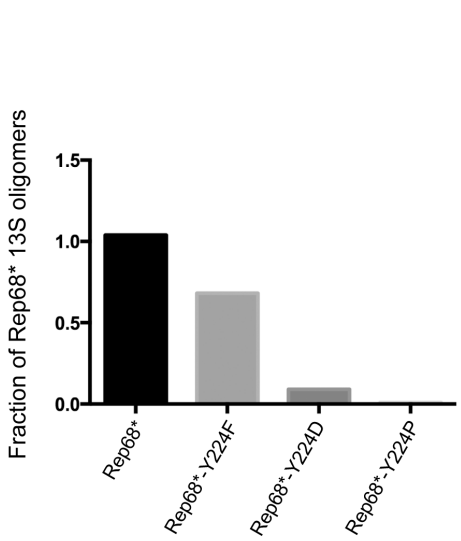
Table 1: Rep68 and interface mutants binding constants on *AAVS1* and p5 RBS-containing DNA.

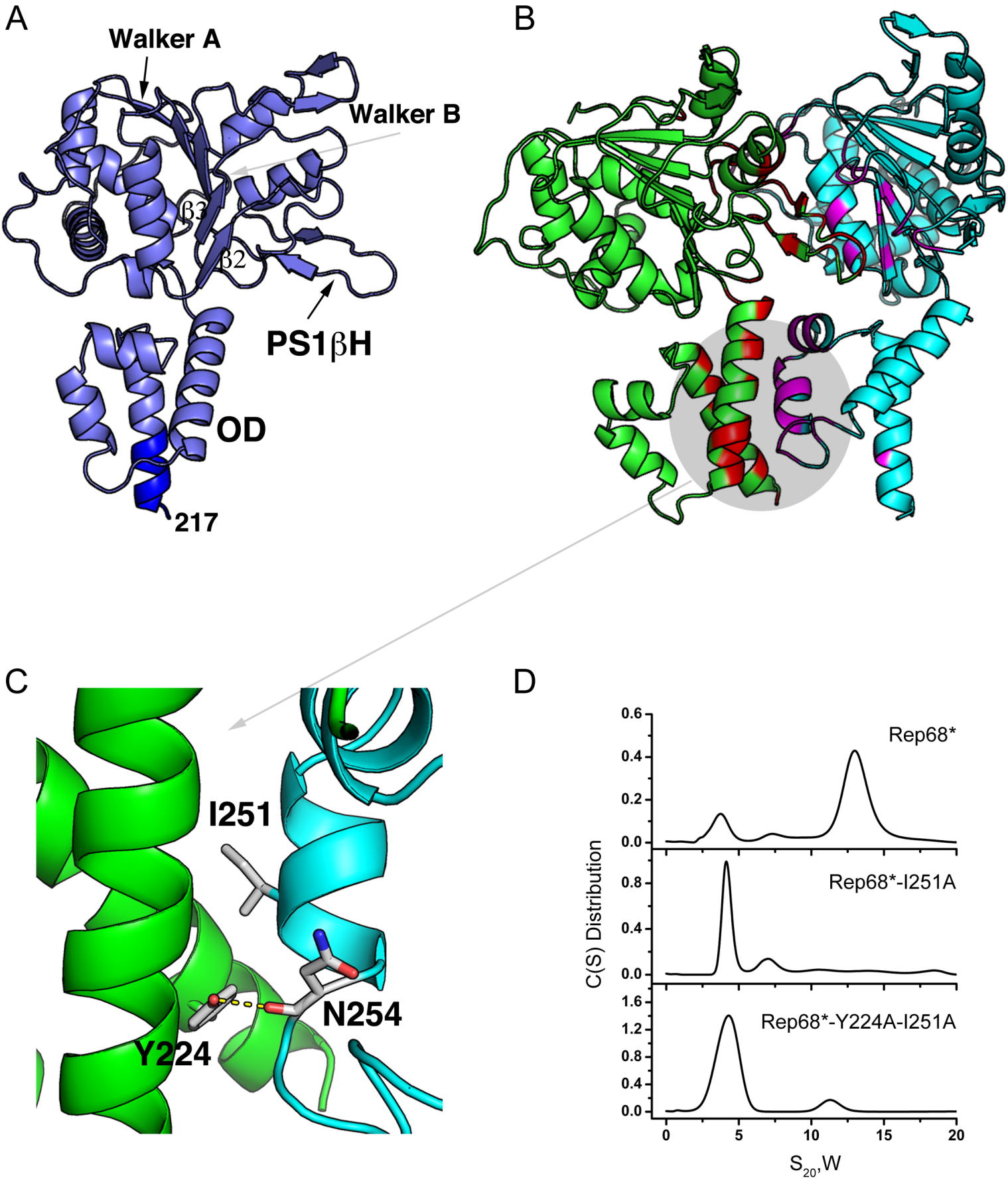
Protein	*AAVS1*-41 (nM)	p5-41 (nM)
Rep68*	128	203
Rep68*-K340H	123	136
Rep68*-Y224A	nd*	nd*
Rep68*-Y224F	221	311
Rep68*-I251A	1438	nd*
Rep68*-YI	nd*	nd*

*not determined due to poor binding.

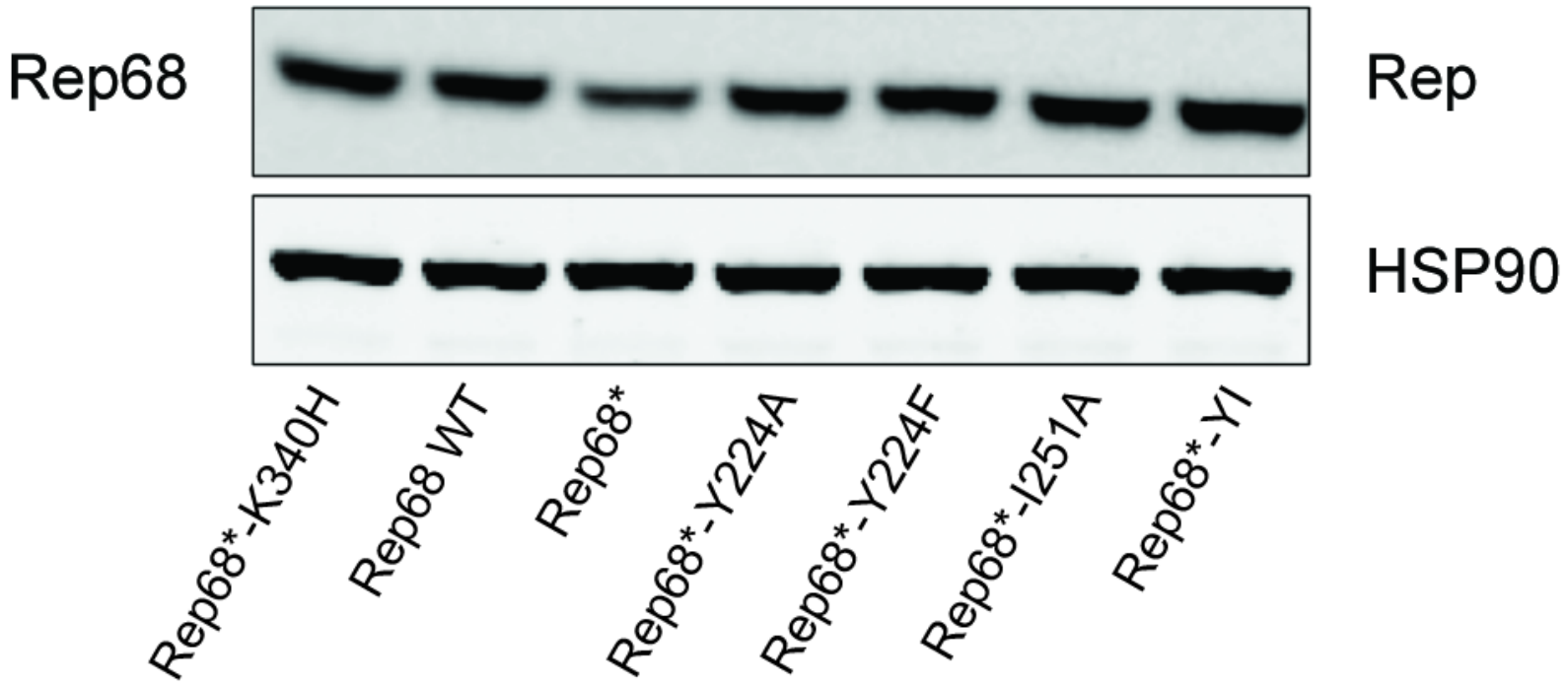

A
1
225
409
OBD
HD
208
206-NQNPNSDAPVIRSKTSARY-224


B
Rep68*
Rep68*-Y224F
Rep68*-Y224P
Rep68*-Y224D
C(S) Distribution
S20,W

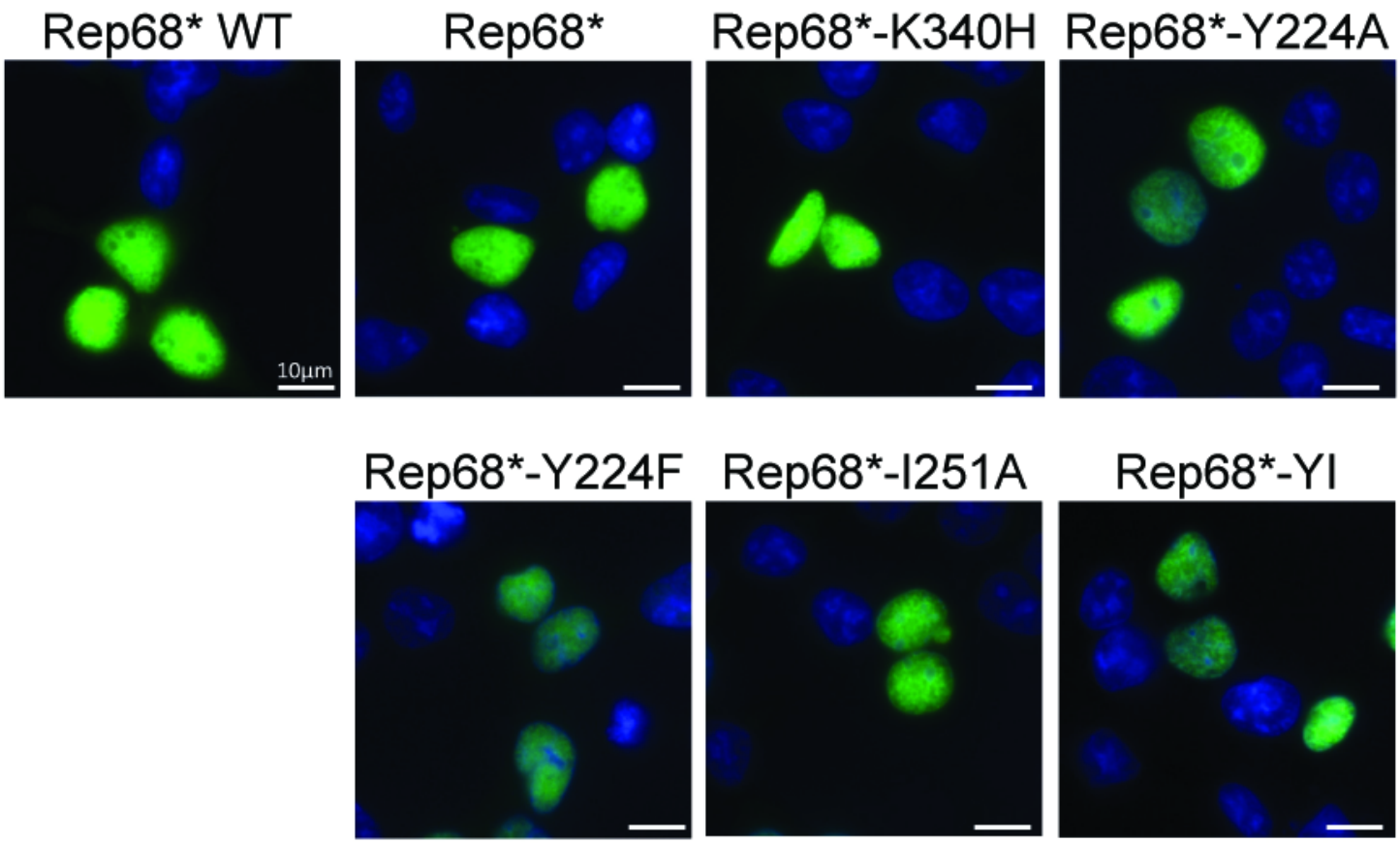

C
Fraction of Rep68* 13S oligomers
Rep68*
Rep68*-Y224F
Rep68*-Y224D
Rep68*-Y224P

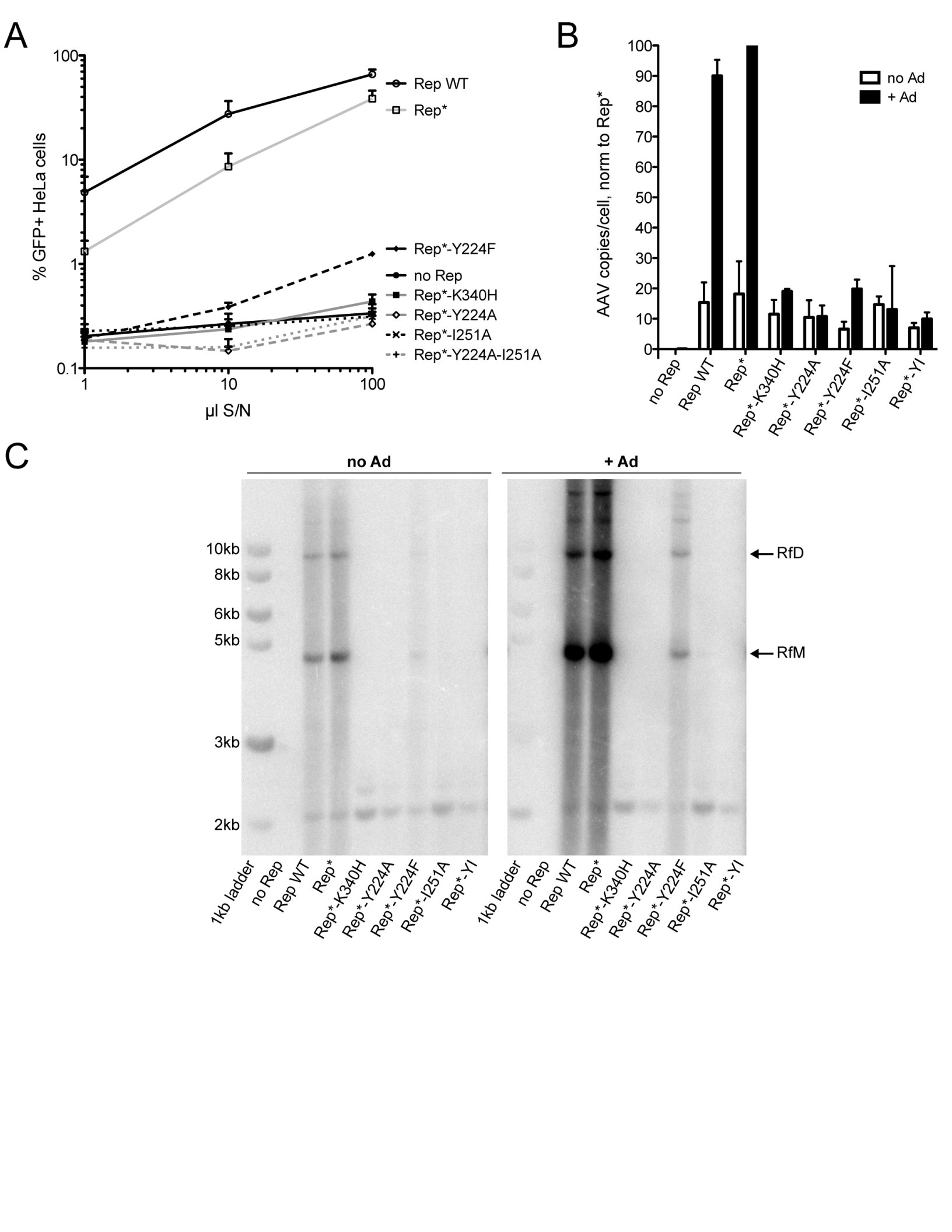

A
Walker A
Walker B
β3
β2
PS1βH
OD
217
B
C
I251
N254
Y224
D
Rep68*
Rep68*-I251A
Rep68*-Y224A-I251A
C(S) Distribution
S20,W
0.6
0.4
0.2
0.0
0.8
0.4
0.0
1.6
1.2
0.8
0.4
0.0
0
5
10
15
20

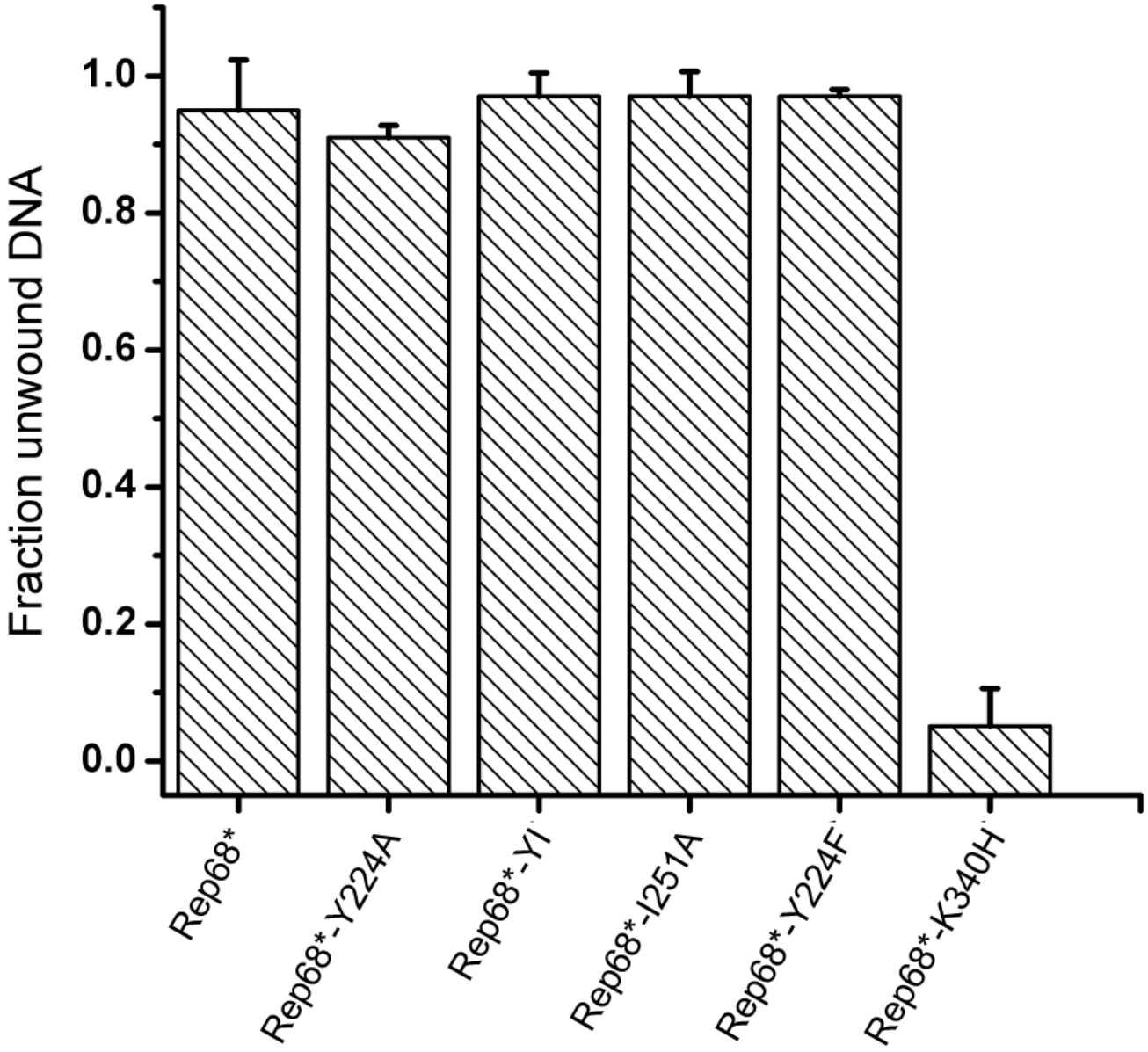
A

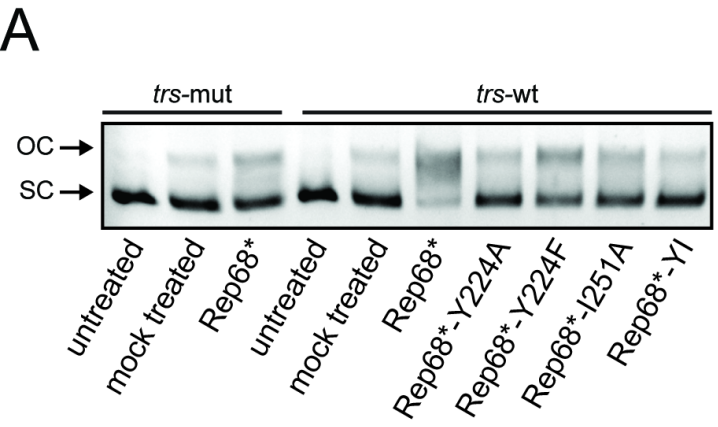

Rep68
Rep
HSP90
Rep68*-K340H
Rep68 WT
Rep68*
Rep68*-Y224A
Rep68*-Y224F
Rep68*-I251A
Rep68*-YI

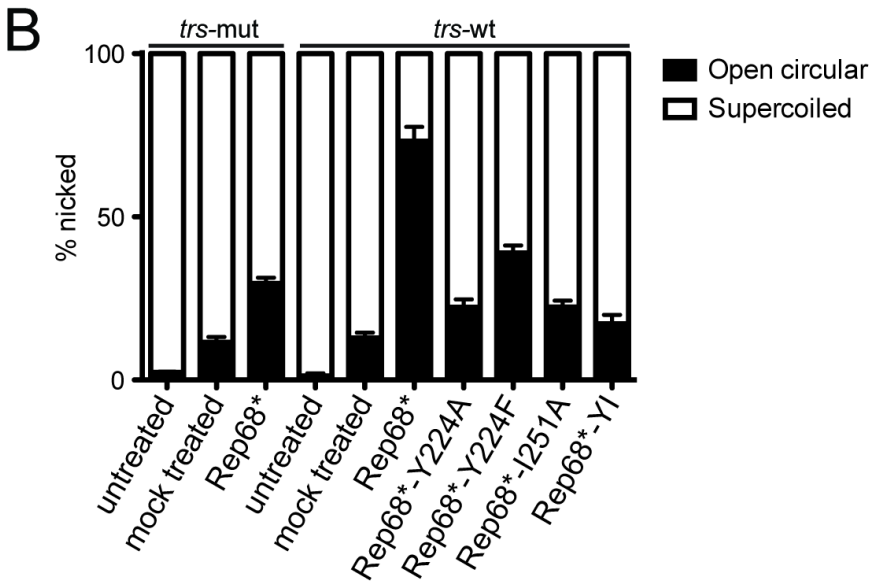
B

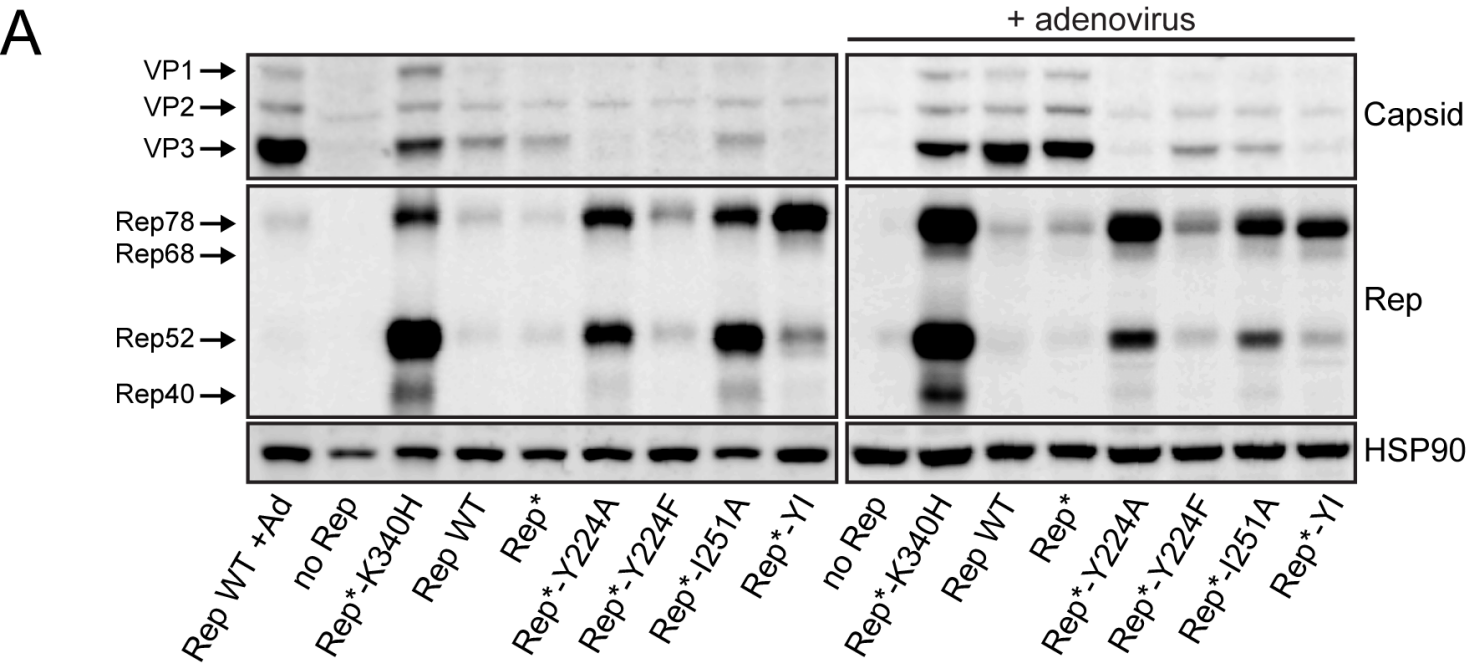

Rep68* WT
Rep68*
Rep68*-K340H
Rep68*-Y224A
10μm
Rep68*-Y224F
Rep68*-I251A
Rep68*-YI

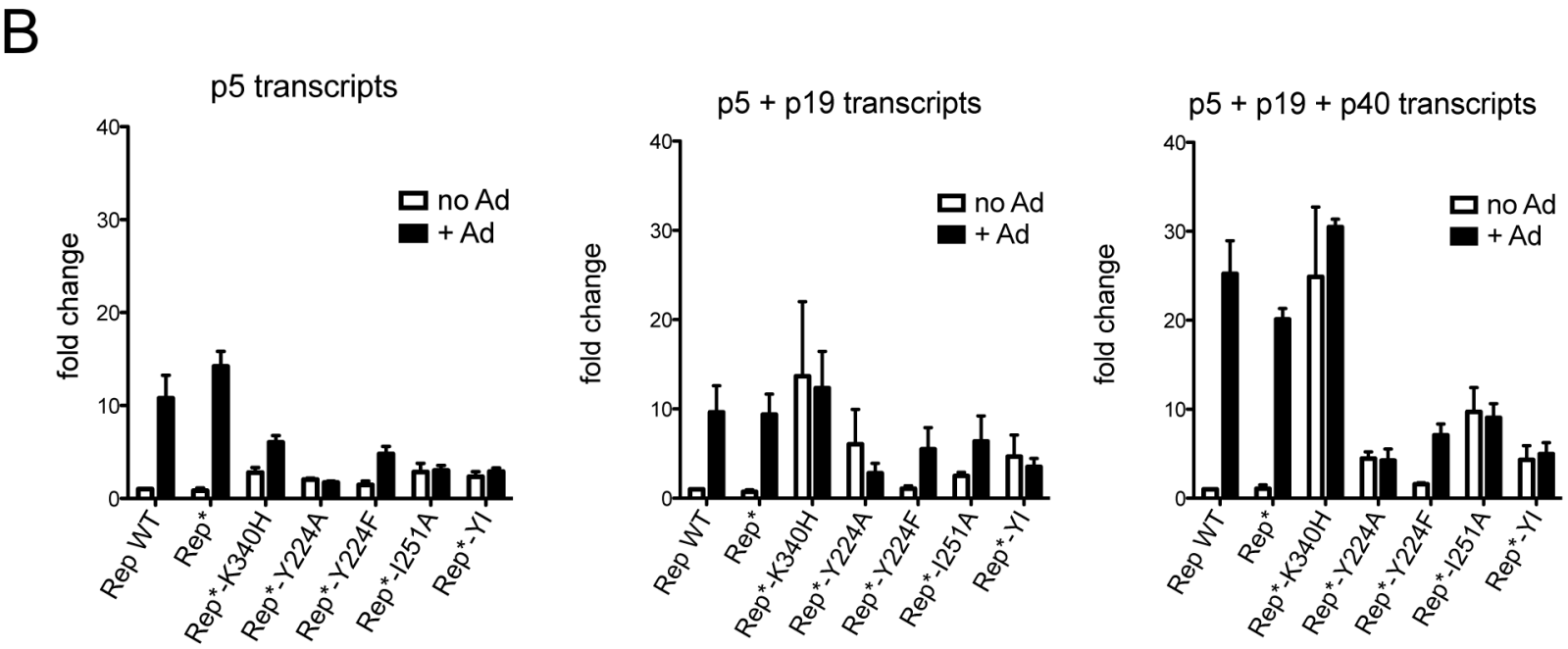

A
100
10
1
0.1
% GFP+ HeLa cells
1
10
100
µl S/N
Rep WT
Rep*
Rep*-Y224F
no Rep
Rep*-K340H
Rep*-Y224A
Rep*-I251A
Rep*-Y224A-I251A
B
AAV copies/cell, norm to Rep*
100
90
80
70
60
50
40
30
20
10
0
no Ad
+ Ad
no Rep
Rep WT
Rep*
Rep*-K340H
Rep*-Y224A
Rep*-Y224F
Rep*-I251A
Rep*-YI
C
no Ad
+ Ad
10kb
8kb
6kb
5kb
3kb
2kb
RfD
RfM
1kb ladder
no Rep
Rep WT
Rep*
Rep*-K340H
Rep*-Y224A
Rep*-Y224F
Rep*-I251A
Rep*-YI


Fraction unwound DNA
1.0
0.8
0.6
0.4
0.2
0.0
Rep68*
Rep68*-Y224A
Rep68*-Y1
Rep68*-I251A
Rep68*-Y224F
Rep68*-K340H


A
trs-mut
trs-wt
OC
SC
untreated
mock treated
Rep68*
untreated
mock treated
Rep68*
Rep68*-Y224A
Rep68*-Y224F
Rep68*-I251A
Rep68*-YI


B
trs-mut
trs-wt
100
50
0
% nicked
Open circular
Supercoiled
untreated
mock treated
Rep68*
untreated
mock treated
Rep68*
Rep68*-Y224A
Rep68*-Y224F
Rep68*-I251A
Rep68*-YI


A
+ adenovirus
VP1
VP2
VP3
Capsid
Rep78
Rep68
Rep52
Rep40
Rep
HSP90
Rep WT +Ad
no Rep
Rep*-K340H
Rep WT
Rep*
Rep*-Y224A
Rep*-Y224F
Rep*-I251A
Rep*-YI
no Rep
Rep*-K340H
Rep WT
Rep*
Rep*-Y224A
Rep*-Y224F
Rep*-I251A
Rep*-YI


B
p5 transcripts
p5 + p19 transcripts
p5 + p19 + p40 transcripts
fold change
no Ad
+ Ad
0
10
20
30
40
Rep WT
Rep*
Rep*-K340H
Rep*-Y224A
Rep*-Y224F
Rep*-I251A
Rep*-YI